Identification and inference with ranking restrictions

Pooyan Amir-Ahmadi
Department of Economics, University of Illinois at Urbana-Champaign

Thorsten Drautzburg
Research Department, Federal Reserve Bank of Philadelphia

We propose to add ranking restrictions on impulse-responses to sign restrictions to narrow the identified set in vector autoregressions (VARs). Ranking restrictions come from micro data on heterogeneous industries in VARs, bounds on elasticities, or restrictions on dynamics. Using both a fully Bayesian conditional uniform prior and prior-robust inference, we show that these restrictions help to identify productivity news shocks in the data. In the prior-robust paradigm, ranking restrictions, but not sign restrictions alone, imply that news shocks raise output temporarily, but significantly. This holds both in an application with rankings in the form of heterogeneity restrictions and in another applications with slope restrictions as rankings. Ranking restrictions also narrow bounds on variance decompositions. For example, the bound of the contribution of news shocks to the forecast error variance of output narrows by about 30 pp at the one-year horizon. While misspecification can be a concern with added restrictions, they are consistent with the data in our applications.

Keywords. Structural VAR, set-identification, sign restrictions, ranking restrictions, heterogeneity, posterior bounds, Bayesian inference, sampling methods, productivity news.

JEL classification. C32, C53, E32.

1. Introduction

While SVARs have remained the workhorse for analyzing the dynamic effects of macroeconomic shocks since Sims (1980), the standards for identification have risen: Much of the literature has abandoned the traditional zero restrictions on short-run or long-run responses. One popular strand of the literature, going back to Uhlig (2005), Faust (1998), and Canova and De Nicolo (2002), abandons point identification. Most prominently, Uhlig (2005) introduced qualitative sign restrictions on impulse response functions to

identify a set of impulse response functions, or other structural parameters. Several recent papers, both frequentist and Bayesian, have proposed inference over the parameters in the identified set, or the identified set directly (Granziera, Moon, and Schorfheide (2018), Gafarov, Meier, and Montiel Olea (2018, 2016), Giacomini and Kitagawa (2018)). We build on this approach to set identification and show how to use rankings of impulse-responses to sharpen identification.

Researchers often have qualitative prior information beyond the sign of impulse-responses that can be used for identification. De Graeve and Karas (2014) used "heterogeneity restrictions" that rank the responses of banks with and without deposit insurance. Kilian and Murphy (2012) introduced bounds on macroeconomic elasticities such as the oil supply elasticity. Both types of restrictions rank the responses of different variables in a VAR at the same horizon and are formally equivalent. One can also restrict shapes of responses, by ranking the responses of the same variable at different horizons. Ranking responses over time can discriminate between news and surprise shocks, for example, or impose other priors on the shape of responses.[1] We refer to all three types of restrictions as ranking restrictions. We view these as underutilized.[2] Thus, we explore ranking restrictions systematically here.

In set identified VARs, identifying assumptions have effects if they restrict covariances of impulse-responses to be at odds with unconditional covariances of forecast errors. The unconditional covariance is a reduced-form object that averages over all shocks. Sign restrictions shrink the identified set if they impose covariances conditional on a shock that are at odds with these unconditional covariances. Heterogeneity restrictions sharpen identification based on linear combinations of covariances. Slope restrictions sharpen identification if they restrict conditional dynamic covariances in ways that are at odds with their unconditional counterparts. Whether ranking restrictions are useful is thus an empirical matter.

We demonstrate empirically that ranking restrictions can meaningfully sharpen inference by analyzing productivity news shocks. Beaudry and Portier (2006) and, more recently, Kurmann and Sims (2017) have argued that productivity news shocks are an important determinant of output fluctuations. In the literature, TFP news shocks raise the value of stocks, are expansionary, and lead to higher TFP in the future. We impose these patterns as sign restrictions and add three different types of ranking restrictions in two closely related applications. In the first application, a nine-variable VAR, we introduce disaggregated data in the form of stock returns for five Fama and French (1997) industries. The industries differ in their R&D intensity. We impose the heterogeneity restrictions that more innovative industries returns load more heavily on TFP news. Motivated by the literature, we also consider a macro elasticity bound, restricting the impact response of productivity following a news shock. Last, we show that without additional slope restrictions, the identified set likely includes both productivity surprises and productivity news shocks. We also analyze how to discriminate between different shocks to productivity in the second application. There, we identify news and surprise shocks to

---

[1]Plagborg-Møller (2019) proposed shape restrictions for point-identified impulse-response function estimation.

[2]For example, Uhlig (2017) advocated elasticity bounds as easy to convey to audiences (p. 101).

TFP in a five-variable VAR, along with a monetary policy shock. We use slope restrictions to distinguish the two productivity shocks: TFP initially rises after a news shock, but is mean-reverting after a surprise shock. Ranking restrictions sharpen inference substantially relative to pure sign restrictions in both applications.

We provide two types of inference. One type is fully Bayesian, which uses a conditionally uniform prior (Uhlig (2017)). This prior uses the Haar measure popularized by Uhlig (2005). Baumeister and Hamilton (2015) criticized such priors because they may induce undesirable priors over structural parameters and are never updated by the data. They propose to parameterize the prior over the structural parameters directly. This alternative approach may be challenging in higher dimensional VARs, which would require priors over many elasticities, or when dynamic restrictions are required (see Kilian and Zhou (2018)). Prior-robust inference, the second type of inference we use, allow us to simultaneously use multihorizon restrictions and prior information on elasticities via bounds. Giacomini and Kitagawa (2018) characterized prior-robust inference with general sign restrictions. Like the earlier work of Faust (1998), this approach provides bounds on structural parameters, which are impulse-response and variance decompositions here. These bounds apply to a family of priors consistent with sign or ranking restrictions. Our approach to prior-robust inference also yields Lagrange multipliers that measure the importance of individual restrictions.

Ranking restrictions yield more precise inference using either the conditionally uniform prior or the prior robust approach. In both applications, ranking restrictions narrow the width of the average 68% robust credible set for the response of output and TFP to productivity news by more than 30%. Similarly, the posterior median reduction in the upper bound on the output response is around 30 pp at the one-year horizon in both applications. With the conditionally uniform prior, we generally also obtain narrower credible sets. In some cases, however, ranking restrictions put more mass on stronger responses with a conditionally uniform prior. This can be associated with wider credible sets when the expected response shifts away from zero. In both applications, Lagrange multipliers confirm that ranking restrictions are more important than standard sign restrictions for sharpening the upper bound. In the smaller VAR with its multiple identified shocks, the Lagrange multipliers also show that the identified set for the responses to news shocks is unaffected by the restrictions on the other identified shocks.[3]

A natural concern about additional restrictions is whether they are consistent with the data. Following Kline and Tamer (2016) and Giacomini, Kitagawa, and Volpicella (2017), we therefore characterize the posterior probability that the identified set is nonempty. Giacomini and Kitagawa (2018) referred to this probability as the posterior plausibility. The added restrictions allow us to compute a conservative test of the restrictions: If the posterior probability of an empty set increases significantly when ranking restrictions are added to the sign restrictions, this provides evidence against the restrictions. With proper priors over the reduced-form VAR parameters, we also report a

---

[3]To draw from the conditionally uniform prior, we develop a new Gibbs sampler, described in the Online Supplementary Material in Appendix A (Amir-Ahmadi and Drautzburg (2021)). Unlike the accept-reject samplers in the literature (see Uhlig (2005), Rubio-Ramírez, Waggoner, and Zha (2010)), this sampler directly draws from the identified set for the underlying rotation matrices.

closely related Bayes factor. To compute the probabilities, we develop a fast computational condition that is necessary and sufficient to determine whether the identified set for a single shock has positive measure. The prior-robust posterior plausibility coincides with the fully Bayesian posterior probability under the conditionally uniform prior.

The qualitative ranking restrictions are consistent with the data in our two applications, based on their posterior plausibility. This result is not mechanical. In our first application, we show that posterior plausibilities quickly decline when strengthening the heterogeneity restrictions beyond the qualitative restrictions we propose. We also show that, combined with slope restrictions, the data reject a model with reversed heterogeneity restrictions. Such tests are, thus, useful for researchers wishing to assess the plausible strength of restrictions or possible misspecification.

Structure and notation.   First, we set up the general statistical model and identification problem with sign and ranking restrictions. We provide intuition for the determinants of the identified sets in small VARs in the case of heterogeneity and slope restrictions. Second, we show how to detect whether identified sets are nonempty for given restrictions and discuss inference. Third, we analyze productivity news shocks. The Appendix in the Online Supplementary Material (Amir-Ahmadi and Drautzburg (2021)) contains all proofs and additional results.

We use bold lower (upper) case letters to denote vectors (matrices). $\|\mathbf{a}\| = \sqrt{\mathbf{a}'\mathbf{a}}$ is the Euclidean norm. $\mathbf{e}_i$ is a vector of zeros with a one in its ith position. $\boldsymbol{\Sigma}^{\mathrm{tr}}$ denotes the lower triangular Cholesky factor of $\boldsymbol{\Sigma}$. A_{ij} denotes element (i, j) of $\mathbf{A}$ and $a_i = (\mathbf{a})_i$ denotes the ith element of $\mathbf{a}$. $\mathbf{A}'$ denotes the transpose of $\mathbf{A}$. $\otimes$ represents the Kronecker product. $\mathcal{O}_n$ denotes the space of n-dimensional orthonormal matrices. $\mathbb{1}_{\{\circ\}}$ denotes the indicator function. The scalar $\mathfrak{s}$ is in $\{-1, 0, 1\}$.

2. Model

Here, we set up the standard Bayesian VAR framework and define sign and ranking restrictions. Several examples illustrate the concept of ranking restrictions. We then develop intuition about how sign and ranking restrictions affect the identified set in small-scale VARs.

2.1 *Setup*

We consider a Gaussian VAR in the $n \times 1$ vector of observables $\mathbf{y}_t$. The VAR has p lags and *iid* normally distributed forecast errors $\mathbf{u}_t$:

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{B}_i \mathbf{y}_{t-i} + \mathbf{u}_t, \quad \mathbf{u}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2.1}$$

where $\mathbf{c}$ is the constant term, $\mathbf{B}_i$, $i \geq 1$ denotes the lag coefficients, and $\boldsymbol{\Sigma}$ the covariance matrix.

Structural VARs are underidentified and require additional restrictions to map the reduced-form innovations $\mathbf{u}_t$ to structural shocks $\boldsymbol{\varepsilon}_t$ by factoring $\boldsymbol{\Sigma}$. This can be summarized as follows:

$$\mathbf{u}_t = \mathbf{A}\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_n), \qquad \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'. \tag{2.2}$$

Besides this generic VAR restriction on $\mathbf{A}$, we impose qualitative restrictions on the impulse-responses to set-identify the VAR model. We now lay out notation to formalize these restrictions.

We define impulse vectors following Uhlig (2005).

DEFINITION 1 (Impulse vectors). The vectors $\mathbf{a}_s \in \mathbb{R}^n$, $s = 1, \ldots, S$, $S \leq n$ are called *impulse vectors*, iff there is some matrix $\mathbf{A}$ with $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}$ so that each $\mathbf{a}_s$ is a column of $\mathbf{A}$.

Formally, let $\boldsymbol{\Sigma}^{\text{tr}}$ be the lower Cholesky matrix and take any orthonormal matrix $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$. Then the columns of $\boldsymbol{\Sigma}^{\text{tr}}\mathbf{Q}$ are impulse vectors. We can thus express impulse vectors without loss of generality as

$$\mathbf{a}_s = \boldsymbol{\Sigma}^{\text{tr}}\mathbf{q}_s, \qquad \mathbf{q}_s'\mathbf{q}_s = 1, \qquad \mathbf{q}_s'\mathbf{q}_{\bar{s}} = 0, \quad \bar{s} \neq s. \tag{2.3}$$

The orthogonality restriction $\mathbf{q}_s'\mathbf{q}_{\bar{s}} = 0$ matters only if we restrict multiple shocks. For unrestricted shock, we can always construct vectors $\mathbf{q}_{\bar{s}}$ in the Nullspace of the restricted shocks. Generally, the impulse vectors $\mathbf{a}_s$ have no economic interpretation, unless they satisfy economic restrictions introduced below.

Defining $\mathbf{B} \equiv [\mathbf{c}, \mathbf{B}_1, \ldots, \mathbf{B}_p]$, we can now write our full model as

$$p(\mathbf{y}^T, \mathbf{B}, \boldsymbol{\Sigma}, \mathbf{Q}) = \ell(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{y}^T)\pi_0(\mathbf{B}, \boldsymbol{\Sigma})\pi_{\mathbf{Q}}(\mathbf{Q}|\mathbf{B}, \boldsymbol{\Sigma}), \tag{2.4}$$

where $\mathbf{y}^T$ collects the history of observables, ℓ is the likelihood function, π_0 denotes the prior over the identifiable reduced-form parameters, and $\pi_{\mathbf{Q}}$ denotes the prior over $\mathbf{Q}$ that incorporates restrictions on impulse responses. We later assume a standard conjugate prior over $(\mathbf{B}, \boldsymbol{\Sigma})$ and take these parameters for now as given. We discuss estimation and comparison of models with different restrictions on $\mathbf{Q}$ in Section 3. Now, we focus on what the reduced-form parameters and beliefs about impulse responses imply about $\mathbf{Q}$. We assume that $\pi_{\mathbf{Q}}$ has full support over the identified set.

2.1.1 *Sign and ranking restrictions*   To identify structural impulse vectors, we impose qualitative restrictions on the impulse-responses that $\mathbf{Q}$ induces. We need extra notation to define these restrictions. Given the companion form $\mathbf{x}_t = \boldsymbol{\mu} + \mathbf{B}_x\mathbf{x}_{t-1} + \mathbf{A}_x\boldsymbol{\varepsilon}_t$, $\mathbf{x}_t \equiv [\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p+1}']'$ of the VAR (2.1) the vector of impulse-response at horizon h is

$$\mathbf{r}_{\mathbf{a}}^h = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n, n \times (p-1)} \end{bmatrix} (\mathbf{B}_x)^h \begin{bmatrix} \mathbf{a} \\ \mathbf{0}_{n \times (p-1), 1} \end{bmatrix}. \tag{2.5}$$

Restrictions on a single shock.   We now define sign restrictions on a single impulse-vector, following Amir-Ahmadi and Uhlig (2015). Later, we extend the setup to multiple shocks.

Imposing sign restrictions is equivalent to picking a list $\mathbb{L}_{SR} \subseteq \{(\mathfrak{s}, k) | \mathfrak{s} \in \{-1, 0, 1\}, k \in \{1, \ldots, n\}\}$ of variables k and signs $\mathfrak{s}$ as well as a restriction horizon $H \geq 0$. We allow for $\mathfrak{s} = 0$ to accommodate restrictions for a subset of the maximum horizon H. For example, there could be no impact restriction with news shocks.

DEFINITION 2 (Sign restrictions). The impulse vector **a** satisfies the *sign restrictions* $(\mathbb{L}_{SR}, H)$ iff $\mathfrak{s}_h \times \mathbf{r}_{\mathbf{a}, k}^h \geq 0$ for all $(\mathfrak{s}_h, k) \in \mathbb{L}_{SR}$ and $h \in \{0, \ldots, H\}$.

We define ranking restrictions similarly, but for pairs of variables (n, m) and a pair of horizons $(h, h + \bar{h}_\lambda)$. We use $\bar{h}_\lambda \neq 0$ for shape restrictions, as explained below. Ranking restrictions also have an associated strength $\lambda \in \mathbb{R}_+$. Define $\mathbb{L}_{RR} \subseteq \{(\mathfrak{s}, k, \ell, \lambda, \bar{h}_\lambda) | \mathfrak{s} \in \{-1, 0, 1\}, (k, \ell) \in \{1, \ldots, n\}^2, \lambda \geq 0\}$. Ranking restrictions nest sign restrictions if $\lambda = 0$.

DEFINITION 3 (Ranking restrictions). The impulse vector **a** satisfies the *ranking restrictions* $(\mathbb{L}_{RR}, H)$ iff $\mathfrak{s}_h \times r_{a,k}^h \geq \lambda \mathfrak{s}_h \times r_{a,\ell}^{h + \bar{h}_\lambda}$ for all $(\mathfrak{s}, k, \ell, \lambda, \bar{h}_\lambda) \in \mathbb{L}_{RR}$ and $h \in \{0, \ldots, H\}$.

Sign or ranking restrictions shape the identified set of **q** and the corresponding identified set of structural parameters, defined as follows.

DEFINITION 4 (Identified set for **q**). The identified set $\mathcal{Q}(\mathbb{L}, H)$ is the collection of all **q**, with $\mathbf{q}'\mathbf{q} = 1$ for which $\mathbf{a} = \mathbf{\Sigma}^{tr}\mathbf{q}$ satisfies the restrictions in $(\mathbb{L}, H)$.

DEFINITION 5 (Identified set for structural parameters). The identified set $\mathcal{F}(f \mid \mathbb{L}, H) = f(\mathcal{Q}(\mathbb{L}, H))$ is the set of all $f(\mathbf{q})$ with **q** in the identified set for **q**, where $f(\mathbf{q})$ is some objective function.

$f(\mathbf{q})$ can be any function of **q**, such as an impulse response function or a variance decomposition, and is implicitly indexed by the reduced form parameters $\mathbf{A}$, $\mathbf{\Sigma}^{tr}$.

For the identified set, there is a clear sense in which heterogeneity restrictions improve inference compared to sign restrictions: Heterogeneity restrictions can nest the standard sign restrictions. One can easily formalize that if they nest the sign restrictions, the identified set is weakly smaller. Below and in Appendix B, we provide conditions under which the identified sets are also strictly smaller than with pure sign restrictions in small-scale VARs. A specific prior over **Q**, however, can lead to more dispersed credible sets on the interior of the identified set: The additional restrictions can shift mass away from the center of the prior toward the tails of the distribution.[4]

Ranking restrictions may also apply when no sign restrictions are available because we can only sign the difference in the responses. For example, we might know that lump-sum fiscal transfers raise the expenditure of highly leveraged households more than

---

[4]For example, we find that the TFP response has a lower (prior-robust) upper bound with ranking restrictions, but more mass away from zero than with sign restrictions under the conditionally uniform prior.

those with low leverage. Depending on how the transfers are financed, some households might actually cut expenditures, for example, if they pay most taxes. In that case, we might want to impose only heterogeneity restrictions that do not nest the standard sign restrictions.[5]

Restrictions on multiple shocks.   We build on the notation above to deal with multiple identified shocks. The analogues to Definition 4 and Definition 5 for multiple shocks are the following.

DEFINITION 6 (Identified set for **Q**: multiple shocks). The identified set $\mathcal{Q}((\mathbb{L}_s, H_s)_{s=1}^S)$ is the collection of all $[\mathbf{q}_1, \ldots, \mathbf{q}_S]$, with $[\mathbf{q}_1, \ldots, \mathbf{q}_S]'[\mathbf{q}_1, \ldots, \mathbf{q}_S] = \mathbf{I}_S$ for which $\mathbf{a}_s = \mathbf{\Sigma}^{\text{tr}}\mathbf{q}_s$ satisfy the restrictions in $\mathbb{L}_s$, H_s for all $s = 1, \ldots, S$.

DEFINITION 7 (Identified set for structural parameters: multiple shocks). The identified set $\mathcal{F}(f \mid (\mathbb{L}_s, H_s)_{s=1}^S) = f((\mathbb{L}_s, H_s)_{s=1}^S)$ is the set of all $f([\mathbf{q}_1, \ldots, \mathbf{q}_S])$ with $[\mathbf{q}_1, \ldots, \mathbf{q}_S]$ in the identified set $\mathcal{Q}((\mathbb{L}_s, H_s)_{s=1}^S)$, where $f(\cdot)$ is some objective function.

The only link between the restrictions is through the orthogonality condition on the columns of the rotation matrix, and thereby on the shocks. This link may matter more when the number of identified shocks is close to the number of variables in the VAR. For example, if $S = n = 2$, then $\mathbf{q}_2$ is already pinned down from the unit-norm and the orthogonality to $\mathbf{q}_1$ up to sign, without any additional restriction.

2.1.2 *Types of ranking restrictions*   In our applications, we use three types of ranking restrictions: Heterogeneity restrictions on micro responses, restrictions on macro elasticities, and slope restrictions. The former two restrict different variables at the same horizon, while the latter restricts the same variable at different horizons.

Heterogeneity restrictions.   For heterogeneity restrictions, we have a theory of a qualitative ordering of disaggregate responses and simply pick pairs (k, ℓ). We set $\bar{h}_\lambda = 0$, $\lambda = 1$ for the restricted pairs while imposing standard sign restrictions on the variable with the weakest response. As we show below, however, one can also consider $\lambda \neq 1$ and let the data discipline the choice of λ.

Elasticity restrictions.   Beliefs about macroeconomic relationships can generate bounds that translate to heterogeneity restrictions of varying strengths. For example, to restrict impact multipliers for government spending to be smaller than two in a VAR in log-levels, write

$$\mathbf{r}^0_{\mathbf{a},\text{output}} < \lambda \times \mathbf{r}^0_{\mathbf{a},\text{government spending}} \quad \text{with } \lambda = 2 \times \frac{\bar{G}}{\bar{Y}}.$$

---

[5]In simple cases, ranking restrictions can be equivalent to sign restrictions after a change of variables in a VAR. For example, we could run a VAR in durable and and overall consumption and impose the ranking restriction that durable consumption drops more than overall consumption in response to a monetary policy shock. Equivalently, we could run a VAR in overall consumption and the ratio of durables to overall consumption, and impose that the latter falls. However, this approach breaks down with ranking restrictions relative to multiple other variables.

Thus, our framework allows us to use prior information on some elasticities, as advocated by Baumeister and Hamilton (2015), without specifying full priors for all elasticities. In prior work, Kilian and Murphy (2012) introduced elasticity bounds in the context of oil markets. In an earlier version of this paper, we analyze a fiscal application with an upper bound on multipliers.

When we restrict elasticities to be close to zero, we can also impose approximate zero restrictions (i.e., soft zero restrictions). For example, Christiano, Eichenbaum, and Evans (1999) identified monetary policy shocks via zero short-run restrictions, imposing, among other things, that real output cannot respond contemporaneously to monetary policy shocks. Here, we could also impose an analogous, but less dogmatic, soft zero restriction using the following restrictions:[6]

$$-\lambda \times \mathbf{r}^0_{\mathbf{a},\mathrm{FFR}} < \mathbf{r}^0_{\mathbf{a},\mathrm{GDP}} < \lambda \times \mathbf{r}^0_{\mathbf{a},\mathrm{FFR}} \quad \text{with } \lambda \text{ small, for example, } \lambda = 0.01.$$

We can impose the above restrictions at various horizons, including the long-run.

Slope restrictions.   Ranking impulse-responses to the same variable over time yields what we call slope restrictions. In this case, $k = \ell$ and $h_\lambda \neq 0$. Macroeconomists now consider various types of news shocks, beginning with TFP news (Beaudry and Portier (2006)), and including government spending news (Ramey (2011)) and monetary policy news (Campbell, Evans, Fisher, and Justiniano (2012)). Plagborg-Møller (2019) estimated point-identified IRFs with news shocks directly using a prior on shapes. We use ranking restrictions to condition on shocks that lead to a TFP build-up for TFP news shocks (the first shock) and mean-reversion for TFP surprises:

$$-\lambda \times \mathbf{r}^0_{\mathbf{a}_1,\mathrm{TFP}} > -\mathbf{r}^1_{\mathbf{a}_1,\mathrm{TFP}} > -\lambda \times \mathbf{r}^2_{\mathbf{a}_1,\mathrm{TFP}} \quad \text{with } \lambda = 1,$$
$$\lambda \times \mathbf{r}^0_{\mathbf{a}_2,\mathrm{TFP}} > \mathbf{r}^1_{\mathbf{a}_2,\mathrm{TFP}} > \lambda \times \mathbf{r}^2_{\mathbf{a}_2,\mathrm{TFP}} \quad \text{with } \lambda = 1.$$

These restrictions rule out aliasing of news and surprise shocks. Combining slope restrictions yields other shape restrictions, such as a hump-shapes.

2.2 *Set reduction through ranking restrictions: Examples and intuition*

We now discuss when ranking restrictions sharpen the inference compared to pure sign restrictions in small VARs, building on Granziera, Moon, and Schorfheide (2018).[7] We focus on heterogeneity restrictions in a bivariate VAR with impact restrictions, and summarize extensions. Together, these examples yield intuition about how sign and ranking restrictions impact the identified set.

---

[6]Special care is required for the limit-point of actual zero restrictions: See Arias, Rubio-Ramirez, and Waggoner (2018), who analyze sign and hard zero restrictions in a fully Bayesian fashion using an accept-reject sampler.

[7]Uhlig (2017) also analyzed sign restrictions in a bivariate VAR. He discusses that restrictions on both shocks sometimes, but not always, sharpen identification.

Consider a bivariate VAR and restrict the first shock. Re-write the restrictions using (2.3):

| Standard sign restrictions | Heterogeneity restrictions |

$$\left(\mathbf{r_a^0}\right)_1 \geq 0 \Leftrightarrow q_1 \Sigma_{1,1}^{\mathrm{tr}} \geq 0, \qquad\qquad \left(\mathbf{r_a^0}\right)_1 \geq 0 \Leftrightarrow q_1 \Sigma_{1,1}^{\mathrm{tr}} \geq 0,$$
$$\tag{2.6a}$$

$$\left(\mathbf{r_a^0}\right)_2 \geq 0 \Leftrightarrow q_1 \Sigma_{2,1}^{\mathrm{tr}} + q_2 \Sigma_{2,2}^{\mathrm{tr}} \geq 0, \qquad \left(\mathbf{r_a^0}\right)_2 \geq \lambda\left(\mathbf{r_a^0}\right)_1 \Leftrightarrow \left(q_1 \Sigma_{2,1}^{\mathrm{tr}} + q_2 \Sigma_{2,2}^{\mathrm{tr}}\right) - \lambda q_1 \Sigma_{1,1}^{\mathrm{tr}} \geq 0.$$
$$\tag{2.6b}$$

Since the heterogeneity restriction nests the standard sign restriction for $\lambda = 0$, we proceed by analyzing just the heterogeneity restrictions.[8] See Appendix B for details.

It is easiest to analyze these restrictions graphically. The restrictions form two hyperplanes in $\mathbb{R}^2$. The intersection of these hyperplanes with the unit circle yields the identified set $\mathcal{Q}$. Figure 1, an extension of Granziera, Moon, and Schorfheide (2018), shows the results. Since variances are strictly positive, the first restrictions simply requires q_1 to be weakly positive. The second restriction requires q_2 to lie above a line through the origin, whose slope depends on the covariances of the reduced form forecast errors and λ. Its slope is increasing in λ. Note that the lower-left element of the Cholesky decomposition $\Sigma_{2,1}^{\mathrm{tr}}$ is proportional to the reduced form covariance $\Sigma_{2,1}$.

(a) $\lambda\Sigma_{1,1}^{\mathrm{tr}} - \Sigma_{2,1}^{\mathrm{tr}} \leq 0, \Sigma_{2,1}^{\mathrm{tr}} \geq 0$ (b) $\lambda\Sigma_{1,1}^{\mathrm{tr}} - \Sigma_{2,1}^{\mathrm{tr}} \geq 0, \Sigma_{2,1}^{\mathrm{tr}} \geq 0$ (c) $\lambda\Sigma_{1,1}^{\mathrm{tr}} - \Sigma_{2,1}^{\mathrm{tr}} \geq 0, \Sigma_{2,1}^{\mathrm{tr}} \leq 0$
Weak heterogeneity restriction   Strong heterogeneity restriction   Strong heterogeneity restriction
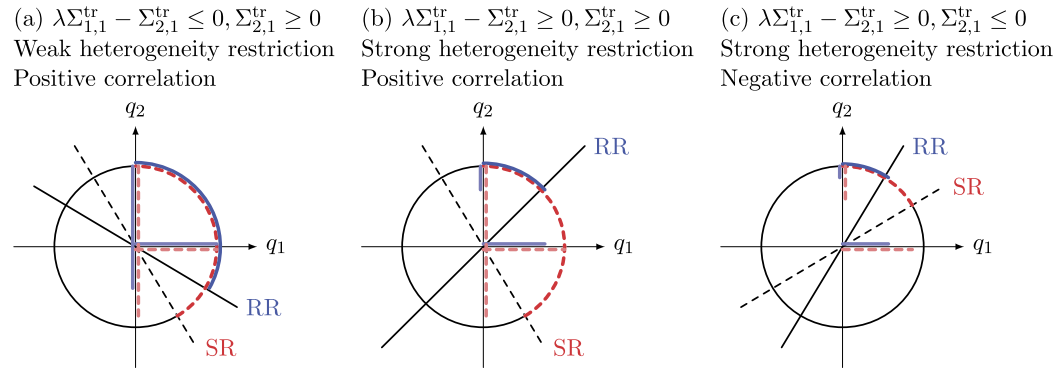Positive correlation   Positive correlation   Negative correlation

FIGURE 1. The identified set with sign and heterogeneity restrictions. Note: The identified set is given by the intersection of the unit circle with the $q_1 > 0$ plane and the plane above the ranking restrictions (RR, here as heterogeneity restrictions) and sign restrictions (SR) lines, respectively. The resultant joint set on the unit circle as well as the marginal sets on the axes are marked in solid lines for the case of RR and in dashed and dashed lines for the traditional SR. The RR set is strictly smaller on the unit circle; this always translates into a tighter set for q_2 and in cases (b) and (c) also in smaller sets for q_1.

---

[8] An example of a VAR with one sign and one heterogeneity restriction is identifying a cost shock in a competitive industry with decreasing returns for which we observe prices and quantities. The restriction that demand is elastic translates to the heterogeneity restriction that minus the quantities fall more than the prices within that industry.

Three cases can arise, given $\lambda > 0$,: (a) the reduced-form covariance is positive and dominates the positive contribution of the heterogeneity restriction, (b) the reduced-form covariance is positive, but the contribution from the heterogeneity restriction dominates, or (c) the covariance is negative so that both contributions are positive. In cases (b) and (c), the identified set for q_1 is strictly smaller with heterogeneity restrictions. In case (a), the identified set for q_1 is $[0, 1]$ in both cases, but the identified set for q_2 is strictly smaller with heterogeneity restrictions.

We are typically interested in the identified set $\mathcal{F}$ for structural parameters such as IRFs rather than in $\mathcal{Q}$. Since $a_1 \propto q_1$, results for the $\mathcal{Q}$ carry over directly to the identified set for the impact IRF of the first variable. For $a_2 = \mathbf{e}_2 \mathbf{\Sigma}^{\text{tr}} \mathbf{q}$, we carry out the extra calculations in B.1.

Intuitively, we find set reductions with sign restrictions if the reduced-form covariance between the variables is of the opposite sign than the one attributed to the identified shock: In this case, the identified shock cannot account for the entire impact response or else the VAR could not generate the observed reduced-form covariance (see Proposition B.1 in Appendix B.1.). This intuition also applies to the case of heterogeneity restrictions, with the reduced-form covariance between the linear combinations $[1, 0]\mathbf{y}_t$ and $[-\lambda, 1]\mathbf{y}_t$ replacing the covariance between variables 1 and 2.

Our analysis implies that for λ large enough, the identified sets for both responses a_1 and a_2 are strictly smaller. A different way to understand our results is through Proposition 4 in Amir-Ahmadi and Uhlig (2015). They show that in a bivariate VAR, all possible sign restrictions are spanned by two sign restrictions with a maximal 180° angle. Standard sign restrictions as defined above imply an angle of 90°, but heterogeneity restrictions imply an angle of more than 90°.[9] As $\lambda \nearrow \infty$, the angle spanned by the heterogeneity restriction approaches 180°. In this case, our identified sets for a_2 converge to a point mass at Σ_{22}^{tr}. This case is informative when we impose a soft zero restriction: For large λ, we are constraining the response of variable one, that is, $(\mathbf{r}_\mathbf{a}^0)_1$, to lie in $[0, \lambda^{-1}(\mathbf{r}_\mathbf{a}^0)_2]$. Given that $(\mathbf{r}_\mathbf{a}^0)_2 > 0$, the limit of $\lambda \nearrow \infty$ is point identification.[10]

Misspecification.   Sign and ranking restrictions can be misspecified, leading to identified sets that exclude the truth. Consider a simple example of a lower triangular VAR—consistent, for example, with a "fast-moving" first variable and a "slow-moving" second variable. Formally, let

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix},$$

---

[9]Because $[1, 0][-\lambda, 1]' < 0$ but $[1, 0][0, 1] = 0$, the angle implied by sign restrictions is wider.

[10]The idea of ranking the responses of two different variables to one shock carries over to ranking the response of a single variable to two different shocks: The response of the first variable to the two shocks can be written as $a_{1,1}(\mathbf{Q}) = \Sigma_{11}^{\text{tr}}[Q_{1,1}, Q_{1,2}]$ subject to $\|[Q_{1,1}, Q_{1,2}]\| = 1$. Assuming positive responses, the heterogeneity restriction then takes the form of $Q_{1,1} > 0$ and $\lambda Q_{1,1} > Q_{1,2} = \sqrt{1 - Q_{1,1}^2} > 0$ so that $Q_{11} > \frac{1}{\sqrt{1+\lambda^2}} > 0$. Because $Q_{12} = \sqrt{1 - Q_{11}^2} = \frac{|\lambda|}{\sqrt{1+\lambda^2}} > 0$, we have a strict set reduction.

where we normalize $a_{11}, a_{22} > 0$. Here,

$$\boldsymbol{\Sigma}^{\text{tr}} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix},$$

so that the true $\mathbf{q}^*$ is $[1, 0]$. Clearly, from case (c) in Figure 1, if $a_{21} < 0$, both sign and ranking restrictions falsely exclude the truth from $\mathcal{Q}$. In contrast, when $a_{21} > 0$, sign restrictions are never misspecified, but ranking restrictions are misspecified if $\lambda > \frac{a_{21}}{a_{11}}$. If $\lambda \in (0, \frac{a_{21}}{a_{11}}]$ the identified set $\mathcal{Q}$ shrinks and still includes $\mathbf{q}^*$. Ranking restrictions can thus narrow the identified set. However, both sign and ranking restrictions can narrow the identified set to exclude the truth. To address this issue, we discuss tests of the restrictions in the next section, albeit conservative ones. In addition, we implement a placebo-type test in our empirical application of heterogeneity restrictions.

Extensions.   The bivariate case with static restrictions is useful to build intuition. However, it does not allow us to speak to the effect ranking restrictions have on responses that are not directly restricted, or to the role of dynamics. In Appendix B, we first cover trivariate VARs and show that a similar condition on covariances as in the bivariate case is sufficient for set reduction for the identified set for the first impact-response $\mathbf{a}_1$. This identified set can be strictly smaller with ranking restrictions involving only the other two variables. Second, we consider dynamic slope restrictions in a bivariate VAR(1). The key intuition is again that the restrictions sharpen inference if they require covariances between IRFs different from the unconditional covariances in the data. With dynamic restrictions, we also provide examples when the identified set is empty. This occurs, for example, if the VAR is mean reverting and the restrictions impose enough persistence.

3. Estimation

The identified set depends on reduced form parameters and rotation matrices. We first summarize the (standard) Bayesian inference on the reduced form parameters. Second, we develop a simple computational criterion for determining whether the identified set has positive measure for given reduced form parameters. This lends itself readily to testing the identifying assumptions relative to VARs with weaker (or fewer) restrictions, the third part of this section. Last, we discuss two types of inference: a Bayesian prior-robust approach and a conditionally uniform prior over $\mathbf{Q}$.

3.1 *Reduced form parameter uncertainty*

We quantify the uncertainty about the reduced form parameters using a Bayesian approach with a natural conjugate prior. The posterior distribution is standard for our Gaussian Bayesian VAR. Defining $\boldsymbol{\beta} \equiv \text{vec}(\mathbf{B})$, we have the following conjugate prior distribution over the reduced form parameters $\mathbf{B}$, $\boldsymbol{\Sigma}$:

$$\boldsymbol{\beta} | \boldsymbol{\Sigma}^{-1} \sim \mathcal{N}\big(\bar{\boldsymbol{\beta}}_0, \mathbf{N}_0^{-1} \otimes \boldsymbol{\Sigma}\big), \tag{3.1}$$

$$\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}_n\big(\nu_0(\bar{\boldsymbol{\Sigma}}_0)^{-1}, \nu_0\big). \tag{3.2}$$

The posterior distribution for $\boldsymbol{\Sigma}^{-1}$ is a Wishart distribution, from which we draw directly. Given the draw for $\boldsymbol{\Sigma}^{-1}$, we can draw from the conditional normal distribution for the coefficients $\boldsymbol{\beta}$. The resultant draws are independent realizations of the reduced form posterior; see Uhlig (1994). We drop reduced form draws if their companion form $\mathbf{B}_x$ has a maximum absolute eigenvalue of ≥ 1.03 or if their identified set $\mathcal{Q}$ is empty.

For future reference, we now map the restrictions $\mathbb{L}_{\mathrm{RR}}$ and the reduced form draws into matrices that encode linear inequality restrictions. For all $j \in \{1, \ldots, J_s\}$ and $h \in \{0, \ldots, H_s\}$,

$$\mathfrak{s}_h^{(j)} \mathbf{e}_{n^{(j)}} \mathbf{r}_\mathbf{a}^h \mathbf{q}_s \geq \mathfrak{s}_h^{(j)} \lambda^{(j)} \mathbf{e}_{m^{(j)}} \mathbf{r}_\mathbf{a}^{h_\lambda} \mathbf{q}_s \quad \Leftrightarrow \quad \underbrace{\left(\mathfrak{s}_h^{(j)} \mathbf{e}_{n^{(j)}} \mathbf{r}_\mathbf{a}^h - \mathfrak{s}_h^{(j)} \lambda^{(j)} \mathbf{e}_{m^{(j)}} \mathbf{r}_\mathbf{a}^{h_\lambda}\right)}_{\equiv \mathbf{w}'_{s,h,j}} \mathbf{q}_s \geq 0. \tag{3.3}$$

Stacking over all J_s restrictions at horizon h, we obtain the $J_s \times n$ dimensional matrix $\tilde{\mathbf{W}}_{s,h} = [\mathbf{w}'_{s,h,j}]_{j=1}^{J_s}$. Stacking over all $h = 0, \ldots, H_s$, we obtain the $(H_s + 1)J_s \times n$ dimensional matrix $\tilde{\mathbf{W}}_s = [\tilde{\mathbf{W}}_{s,h}]_{h=0}^{H_s}$. We define $\mathbf{W}_s$ as the subset of rows of $\tilde{\mathbf{W}}_s$ that are nonzero. This is relevant when $\mathfrak{s}_h^{(j)} = 0$ implies restrictions that apply only to a subset of the horizon $0, \ldots, H_s$. For restrictions on cumulative responses, for example, when the VAR is estimated in differences, replace $\mathbf{r}_\mathbf{a}^h$ with $\sum_{\ell=0}^h \mathbf{r}_\mathbf{a}^h$.

When there are only sign and ranking restrictions involving some common horizons $h = \underline{h}_s, \ldots, H_s$, we can also construct the matrix $\mathbf{W}_s$ directly as

$$\mathbf{W}_s \equiv \begin{bmatrix} \mathbf{S}_s(\mathbf{E}_s - \boldsymbol{\Lambda}_s)\mathbf{B}_x^{\underline{h}_s} \begin{bmatrix} \boldsymbol{\Sigma}^{\mathrm{tr}} \\ \mathbf{0}_{(p-1)\times n} \end{bmatrix} \\ \vdots \\ \mathbf{S}_s(\mathbf{E}_s - \boldsymbol{\Lambda}_s)\mathbf{B}_x^{H_s} \begin{bmatrix} \boldsymbol{\Sigma}^{\mathrm{tr}} \\ \mathbf{0}_{(p-1)\times n} \end{bmatrix} \end{bmatrix}. \tag{3.4}$$

When the restrictions are on cumulative responses, replace $\mathbf{B}_x^h$ with $\sum_{s=0}^h \mathbf{B}_x^s$. The matrix $\mathbf{E}_s$ selects response variables, $(\mathbf{E}_s)_{k\ell} = \mathbb{1}\{n_s^{(k)} = \ell\}$, the matrix $\boldsymbol{\Lambda}_s$ is zero except for rows k with $\lambda^{(k)} \neq 0$, in which case $(\boldsymbol{\Lambda}_s)_{k\ell} = \lambda_s^{(k)} \mathbb{1}\{m_s^{(k)} = \ell\}$, and $\mathbf{S}_s = \mathrm{diag}(\mathfrak{s}_s^{(j)})$ encodes the signs. With $p = 1$ and impact restrictions only, (3.4) simplifies to $\mathbf{W}_s = \mathbf{S}_s(\mathbf{E}_s - \boldsymbol{\Lambda}_s)\boldsymbol{\Sigma}^{\mathrm{tr}}$.

For identified shocks $s = 1, \ldots, S$, we can summarize the identifying restrictions, shock normalizations, and the shock orthogonality conditions as follows:

$$\mathbf{W}_s \mathbf{q}_s \geq 0, \qquad \mathbf{q}'_s \mathbf{q}_s = 1, \qquad \mathbf{q}'_s \mathbf{q}_u = 0, \quad u \neq s. \tag{3.5}$$

Since $\mathbf{W}_s$ depends on the random parameters $\mathbf{B}$ (via $\mathbf{B}_x$) and $\boldsymbol{\Sigma}$ (via $\boldsymbol{\Sigma}^{\mathrm{tr}}$), each reduced form draw d is associated with a draw for $\mathbf{W}_s^{(d)}$. We denote the matrix-valued function constructed above as $\mathbf{W}_{f,s}(\mathbf{B}, \boldsymbol{\Sigma}; \mathbb{L}_{\mathrm{RR}})$, so that $\mathbf{W}_s^{(d)} = \mathbf{W}_{f,s}(\mathbf{B}^{(d)}, \boldsymbol{\Sigma}^{(d)}; \mathbb{L}_{\mathrm{RR}})$.

3.2 *Empty identified set with a single identified shock*

We provide a novel condition for assessing whether identified sets have positive measure: We check that the Chebychev center of the constrained set (prior to normaliza-

tion) is nondegenerate. Intuitively, the Chebychev center is the center of the largest ball inscribed in the constrained set. We prove below that for continuous prior distributions, existence of a Chebychev center $\mathbf{z}_c$ with a ball of radius $r > 0$ around it is equivalent to an identified set with positive measure. To ensure that the problem is well-defined, we additionally restrict the solution to $[-1, 1]^n$, the unit n-cube. If $r > 0$, we can then construct a set of positive measure on the constrained unit n-sphere. And if the identified set has positive measure, some element of it lies strictly in the interior. We can then use that element of construct a candidate Chebychev center that has positive radius.

DEFINITION 8 (Chebychev center). The Chebychev center of the constrained set $\mathbf{W}\mathbf{z} \leq 0$ is given by $\mathbf{z}_c$ that solves

$$\{\mathbf{z}_c, r\} = \arg\max_{\mathbf{z}_c, r \geq 0} r \quad \text{s.t.} \quad \mathbf{W}(\mathbf{z}_c + \mathbf{u}) \leq 0, \quad \mathbf{z}_c + \mathbf{u} \leq 1, \quad -(\mathbf{z}_c + \mathbf{u}) \leq 1, \quad \forall \mathbf{u} : \|\mathbf{u}\| \leq r,$$

where the inequalities are element by element.

The center point $\mathbf{z}_c$ need not be unique. We illustrate this by computing the Chebychev center for the three examples from Figure 1. Figure 2 shows the results. In all three cases, the Chebychev center is nondegenerate. In the first case, however, it is not unique. Still, the constrained set is convex and the objective linear. Any local maximum r^* is therefore a global maximum.

PROPOSITION 1 (Nonempty identified set $\mathcal{Q}$). *Take any continuous prior π_q over $\mathbf{q}$ with strictly positive support on the unit sphere. The identified set for reduced form parameters $(\mathbf{B}, \mathbf{\Sigma})$ has positive measure under π_q iff the Chebychev center $\mathbf{z}_c$ of the constrained set in $[-1, 1]^n$ satisfies $r > 0$ with strict inequality.*

(a) $\lambda\Sigma_{1,1}^{\mathrm{tr}} - \Sigma_{2,1}^{\mathrm{tr}} \leq 0, \Sigma_{2,1}^{\mathrm{tr}} \geq 0$ (b) $\lambda\Sigma_{1,1}^{\mathrm{tr}} - \Sigma_{2,1}^{\mathrm{tr}} \geq 0, \Sigma_{2,1}^{\mathrm{tr}} \geq 0$ (c) $\lambda\Sigma_{1,1}^{\mathrm{tr}} - \Sigma_{2,1}^{\mathrm{tr}} \geq 0, \Sigma_{2,1}^{\mathrm{tr}} \leq 0$
Chebychev center $\mathbf{z}_c$ not unique Chebychev center $\mathbf{z}_c$ unique Chebychev center $\mathbf{z}_c$ unique
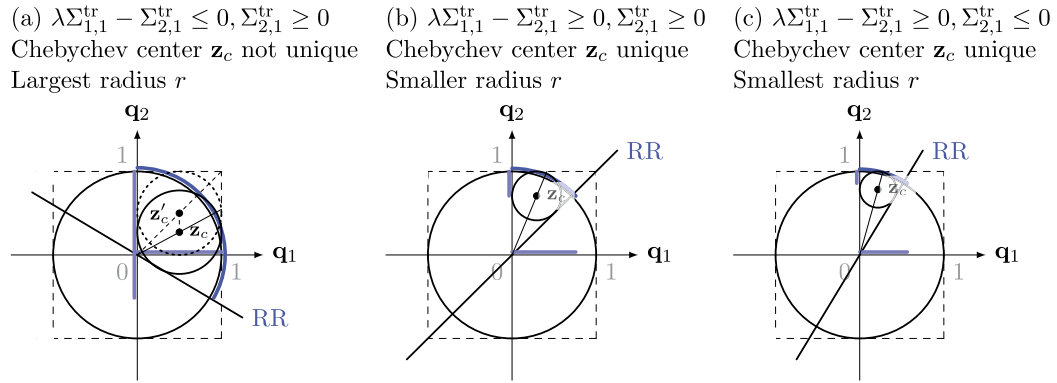Largest radius r Smaller radius r Smallest radius r

FIGURE 2. Chebychev center: Examples in $\mathbb{R}^2$. Note: The plot shows the Chebychev centers $\mathbf{z}_c$ for the cases considered in Figure 1. While the center is not unique in case (a), the radius r of the largest ball inscribed in the intersection of $\mathbf{W}\mathbf{z} \leq 0$ and the unit n-cube is well-defined. If r is strictly positive, the identified set has positive measure under any strictly positive prior on the unit n-sphere.

PROOF.   See Appendix A.1.   □

Our Chebychev center criterion implies the condition for a nonempty identified set in Granziera, Moon, and Schorfheide (2018). In Appendix A.2, we show this formally using Gordan's alternative (Border (2013)). Intuitively, their approach is the dual to ours. They rule out the existence of some nonzero $\mathbf{v} \geq 0$ such that $\mathbf{W}'\mathbf{v} = \mathbf{0}$. If one formulates the search for an interior element of the identified set $\mathcal{Q}$ as a constrained optimization problem, existence of an interior point such as the Chebychev center implies that at these points that the "shadow price" $\mathbf{v}$ of the constraints is zero: Tightening them marginally leaves the optimal solution unaffected. Vice versa, strictly positive shadow prices $\mathbf{v}$ on any of the constraints implies that it is binding with equality, and no interior solution exists.

The key condition for our criterion is whether the radius around the Chebychev center is positive. $\mathbf{z}_c = \mathbf{0}$ is always a candidate solution, since the restrictions always admit a (degenerate) cone through the origin as a solution. As discussed by Granziera, Moon, and Schorfheide (2018) and in Appendix B, multihorizon restrictions can easily cause $\mathcal{Q}$ to be empty. A simple example is the case of persistent restrictions when the VAR displays oscillations. In the case of point identification, the case of $\lambda \nearrow \infty$ in Figure 2(c), the Chebychev center also implies an empty identified set.[11]

3.3 *Testing the identifying restrictions*

Restrictions on impulse-response functions are testable. Piffer (2016) proposed a parametric test. In contrast, we build on Kline and Tamer (2016) to build a test that is free of parametric assumptions about the distribution of $\mathbf{Q}$. Kline and Tamer (2016) proposed, inter alia, the use of posterior model probabilities to assess set-identified models, that is, the probability under the reduced form posterior that $\mathcal{Q}$ is empty. Giacomini and Kitagawa (2018) referred to this quantity as posterior plausibility. We also use this measure, but also report Bayes factors to compare the model that includes ranking restrictions to either the model with only sign restrictions or without any restrictions. These Bayes factors also reflect the prior probability that the identified set is empty. Unlike Giacomini, Kitagawa, and Volpicella (2017), who consider model averaging more generally, we focus on set-identified models. The posterior plausibility and the Bayes factor also apply to our fully Bayesian prior: Because the parametric prior, which we introduce below, is conditionally uniform, it does not change the reduced form posterior density as long as the identified set $\mathcal{Q}$ has positive measure.

The Bayes factor of the model with ranking restrictions (RR) relative to the model with only sign restrictions (SR) relates to the posterior model probabilities as follows,

---

[11]Any $\mathbf{x}_c = [0, c]$, $c > 0$ is a Chebychev center with $\lambda = \infty$. Projected onto the unit circle, this yields a singleton. To admit such zero measure events, one could replace the criterion that $r > 0$ with either $r > 0$ or $\mathbf{x}_c \neq 0$. However, for our fully Bayesian inference we need to rule out point identification.

with $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \boldsymbol{\Sigma})$:

$$\mathrm{BF}_{\mathrm{RR,SR}} \equiv \frac{\displaystyle\int f(\mathbf{y}^T|\boldsymbol{\theta}) \frac{\pi(\boldsymbol{\theta})\mathbb{1}\{\mathrm{RR}\}}{\displaystyle\int \pi(\tilde{\boldsymbol{\theta}})\mathbb{1}\{\mathrm{RR}\}\,d\tilde{\boldsymbol{\theta}}}\,d\boldsymbol{\theta}}{\displaystyle\int f(\mathbf{y}^T|\boldsymbol{\theta}) \frac{\pi(\boldsymbol{\theta})\mathbb{1}\{\mathrm{SR}\}}{\displaystyle\int \pi(\tilde{\boldsymbol{\theta}})\mathbb{1}\{\mathrm{SR}\}\,d\tilde{\boldsymbol{\theta}}}\,d\boldsymbol{\theta}}$$

$$= \frac{\displaystyle\int \mathbb{1}\{\mathrm{RR}\} \frac{f(\mathbf{y}^T|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\displaystyle\int f(\mathbf{y}^T|\tilde{\boldsymbol{\theta}}))\pi(\tilde{\boldsymbol{\theta}}))\,d\tilde{\boldsymbol{\theta}}}\,d\boldsymbol{\theta}}{\displaystyle\int \mathbb{1}\{\mathrm{SR}\} \frac{f(\mathbf{y}^T|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\displaystyle\int f(\mathbf{y}^T|\tilde{\boldsymbol{\theta}})\pi(\tilde{\boldsymbol{\theta}})\,d\tilde{\boldsymbol{\theta}}}\,d\boldsymbol{\theta}} \times \frac{\displaystyle\int \pi(\boldsymbol{\theta})\mathbb{1}\{\mathrm{SR}\}\,d\boldsymbol{\theta}}{\displaystyle\int \pi(\boldsymbol{\theta})\mathbb{1}\{\mathrm{RR}\}\,d\boldsymbol{\theta}}$$

$$= \frac{\displaystyle\int \mathbb{1}\{\mathrm{RR}\}\pi(\boldsymbol{\theta}|\mathbf{y}^T)\,d\boldsymbol{\theta}}{\displaystyle\int \mathbb{1}\{\mathrm{SR}\}\pi(\boldsymbol{\theta}|\mathbf{y}^T)\,d\boldsymbol{\theta}} \times \frac{\displaystyle\int \pi(\boldsymbol{\theta})\mathbb{1}\{\mathrm{SR}\}\,d\boldsymbol{\theta}}{\displaystyle\int \pi(\boldsymbol{\theta})\mathbb{1}\{\mathrm{RR}\}\,d\boldsymbol{\theta}}$$

$$\equiv \frac{\mathbb{E}_T^{\boldsymbol{\theta}}\big[\mathbb{1}\{\mathrm{RR}\}\big]/\mathbb{E}_0^{\boldsymbol{\theta}}\big[\mathbb{1}\{\mathrm{RR}\}\big]}{\mathbb{E}_T^{\boldsymbol{\theta}}\big[\mathbb{1}\{\mathrm{SR}\}\big]/\mathbb{E}_0^{\boldsymbol{\theta}}\big[\mathbb{1}\{\mathrm{SR}\}\big]},$$

where $\mathbb{1}\{\circ\} = 1$ if the identified set $\mathcal{Q}$ has positive measure. The second equality uses that the normalizing constant $\int f(\mathbf{y}^T|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$ of the reduced form posterior $\pi(\boldsymbol{\theta}|\mathbf{y}^T)$ is the same across models. The third equality uses the definition of the reduced form posterior $\pi(\boldsymbol{\theta}|\mathbf{y}^T)$. The last inequality defines expectations over $\boldsymbol{\theta}$, under the reduced form prior and posterior.

The components of the Bayes factor have clear interpretations. The first term, $\frac{\mathbb{E}_T^{\boldsymbol{\theta}}[\mathbb{1}\{\mathrm{RR}\}]}{\mathbb{E}_T^{\boldsymbol{\theta}}[\mathbb{1}\{\mathrm{SR}\}]}$ is the ratio of posterior probabilities of empty sets of Kline and Tamer (2016). Particularly, when computed relative to a model with no sign restrictions or with sign restrictions that are always satisfied, this term is exactly the posterior probability proposed by Kline and Tamer (2016) and from Giacomini and Kitagawa (2018). The second term adjusts for the prior probabilities $\mathbb{E}_0^{\boldsymbol{\theta}}[\mathbb{1}\{\circ\}]$.[12] As noted by Giacomini and Kitagawa (2018), if the ratio $\frac{\mathbb{E}_T^{\boldsymbol{\theta}}[\mathbb{1}\{\mathrm{RR}\}]}{\mathbb{E}_0^{\boldsymbol{\theta}}[\mathbb{1}\{\mathrm{RR}\}]}$ is larger than one, the data favor the plausibility of the restrictions. If the reference model has unitary prior and posterior probabilities, such as models with no restrictions, the Bayes factor only reflects this last ratio.

To implement our test numerically, we use the Chebychev center criterion.

3.4 *Prior-robust inference*

In a standard BVAR with sign restrictions such as Uhlig (2005), the posterior distribution of impulse response functions (IRFs) reflects uncertainty about the reduced form

---

[12]As our derivation shows, $\mathbb{E}_0^{\boldsymbol{\theta}}[\mathbb{1}\{\mathrm{RR}\}]$ is the normalizing constant of the prior.

parameters and the rotation matrix $\mathbf{Q}$. However, the possible prior distributions over $\mathbf{Q}$ may imply different shapes for the resultant IRF, even conditional on reduced form parameters. Baumeister and Hamilton (2015) pointed out that the commonly used prior that $\mathbf{Q}$ is uniformly distributed in the space of orthonormal matrices does not translate to a uniform distribution over the identified set of the structural parameters, since these transform the underlying parameters of the reduced form and the rotation matrices. We also find this in our applications below. Additionally, Arias, Rubio-Ramirez, and Waggoner (2018) argued that practitioners have combined sign and zero restrictions in ways that introduced unnoticed prior information.

One can address the criticism by Baumeister and Hamilton (2015) and side-step Arias, Rubio-Ramirez, and Waggoner (2018) by being conservative and choosing the worst-case prior possible over $\mathbf{Q}$, following Giacomini and Kitagawa (2018). However, the posterior distribution over the reduced-form parameters $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is updated by the data and one can apply standard Bayesian inference to this parameter vector. For each draw of reduced form parameter, we can then compute bounds over the possible distributions of $\mathbf{Q}$.

Specifically, we follow Granziera, Moon, and Schorfheide (2018) to compute the infimum and supremum over all admissible rotation matrices $\mathbf{Q}$. The resulting identified set is distribution-free, as we compute the infimum and supremum over the set of all prior distributions over admissible rotation matrices. We compute this set conditional on the reduced form parameters $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, similar to Faust (1998) and Giacomini and Kitagawa (2018). While the identified set is robust to any full-support prior over rotation matrices, we still care about the parameter uncertainty: Parameter combinations $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ differ in their posterior density. We therefore compute a distribution over the [inf, sup]-bounds that reflects the posterior reduced-form parameter uncertainty.[13]

Formally, define the fully Bayesian posterior distribution over the IRF for variable j at horizon h given the prior π over the rotation vectors q as

$$
\tilde{F}_{j,h}^{\pi}(x) = \int \int_{\mathbf{q}} \mathbb{1}_{\{\mathbf{r}_{\boldsymbol{\Sigma}^{\mathrm{tr}}\mathbf{q}}(h;\boldsymbol{\Sigma},\boldsymbol{\beta})_j \leq x\}} \mathbb{1}_{\{\mathbf{r}_{\boldsymbol{\Sigma}^{\mathrm{tr}}\mathbf{q}}(s;\boldsymbol{\Sigma},\boldsymbol{\beta})_n \geq \lambda \mathbf{r}_{\boldsymbol{\Sigma}^{\mathrm{tr}}\mathbf{q}}(s;\boldsymbol{\Sigma},\boldsymbol{\beta})_m \forall (n,m,\lambda) \in \mathbb{L}_{\mathrm{RR}}^{(j)} \forall j=1,\ldots J\}}
$$
$$
\times \pi(\mathbf{q})\, d\mathbf{q} \times p(\boldsymbol{\Sigma}, \boldsymbol{\beta}; \mathbf{y}^T)\, d\boldsymbol{\Sigma}\, d\boldsymbol{\beta}
$$

In contrast, we define the prior-robust posterior distribution over upper bounds of the IRFs as

$$
F_{j,h}(x) = \int \sup_{\pi,\mathbf{q}|\pi(\mathbf{q})>0} \mathbb{1}_{\{\mathbf{r}_{\boldsymbol{\Sigma}^{\mathrm{tr}}\mathbf{q}}(h;\boldsymbol{\Sigma},\boldsymbol{\beta})_j \leq x\}} \mathbb{1}_{\{\mathbf{r}_{\boldsymbol{\Sigma}^{\mathrm{tr}}\mathbf{q}}(s;\boldsymbol{\Sigma},\boldsymbol{\beta})_n \geq \lambda \mathbf{r}_{\boldsymbol{\Sigma}^{\mathrm{tr}}\mathbf{q}}(s;\boldsymbol{\Sigma},\boldsymbol{\beta})_m \forall (n,m,\lambda) \in \mathbb{L}_{\mathrm{RR}}^{(j)} \forall j=1,\ldots J\}}
$$
$$
\times p(\boldsymbol{\Sigma}, \boldsymbol{\beta}; \mathbf{y}^T)\, d\boldsymbol{\Sigma}\, d\boldsymbol{\beta}
$$

[13] Giacomini and Kitagawa (2018) showed that, under regularity conditions, robust credible sets for IRFs based on such bounds also provide frequentist confidence intervals that are pointwise consistent in level, for any given $(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$. From a frequentist perspective, however, this pointwise consistency is often considered unappealing: Unlike uniform consistency, pointwise consistency does not guarantee adequate size in all samples that are large enough. See Canay and Shaikh (2016) for a discussion.

Algorithm 1 Inference with a single shock.

For $d = 1, \ldots, D$ do:

1. Draw $\mathbf{B}^{(d)}$ and $\mathbf{\Sigma}^{(d)}$ from $p(B, \mathbf{\Sigma}|Y)$.

2. Given $\mathbf{B}^{(d)}$ and $(\mathbf{\Sigma}^{\text{tr}})^{(d)}$, compute $\mathbf{W}^{(d)} = \mathbf{W}_f(\mathbf{B}^{(d)}, \mathbf{\Sigma}; \mathbb{L})$ according to (3.4).

3. Solve for the Chebychev center $\{\mathbf{z}_c, r\}$ of the set $\mathbf{W}^{(d)}\mathbf{z} \leq 0$.

4. Proceed if $r > 0$, that is, the identified set has positive measure. Otherwise, go to Step 1.

   4.(a) *Prior robust variant*: For each variable $i = 1, \ldots, n$ and for each horizon $h = 0, \ldots, H$, calculate bounds:

   $$\min_{\mathbf{q}} \text{ and } \max_{\mathbf{q}} \quad \mathbf{e}'_j(\mathbf{B}_x^{(d)})^h\big[(\mathbf{\Sigma}^{\text{tr}})^{(d)}; \mathbf{0}_{(p-1)n \times n}\big]\mathbf{q}$$

   $$\text{s.t.} \quad \mathbf{W}^{(d)}\mathbf{q} \leq 0, \text{ and } \|\mathbf{q}\| = 1.$$

   Save the resulting values as upper and lower bounds as well as the Lagrange multipliers on the constraints $\mathbf{Wq} \leq 0$.

   4.(b) *Conditionally uniform variant*: Take M draws from the truncated Haar measure $p(\mathbf{q}|\mathbf{B}^{(d)}, \mathbf{\Sigma}^{(d)}; \mathbb{L})$ using either an accept-reject sampler or a direct sampler.

That is, we are evaluating the probability under the best-case prior. The probability that the response of variable j at horizon h is x or lower is, therefore, at most $F_{j,h}(x)$. The definitions for lower bounds and variance decompositions are analogous.

The prior-robust variant of the algorithm we use for inference (Algorithm 1) avoids taking a stance on the shape of the prior over the identified set. It therefore follows the principle of transparent parameterization summarized in Schorfheide (2016). The algorithm combines the standard sampling scheme for the reduced form parameters with the Chebychev center criterion and a numerical optimizer to characterize bounds. This approach mimics our analytical characterization of the identified set in the bivariate VAR examples. It also yields Lagrange multipliers on the constraints as a byproduct of the optimization. These multipliers measure how much restrictions shrink the identified set, as we illustrate in our empirical applications.

The prior-robust variant of Algorithm 1 is similar to the algorithms in Faust (1998) and Giacomini and Kitagawa (2018). Faust (1998) analyzed upper bounds for the variance decomposition. Giacomini and Kitagawa (2018) focused on impulse response functions. The numerical optimization problem in the algorithm has a simple structure: a linear objective, a linear inequality constraint, and an equality constraint with gradient $2\mathbf{q}$. We find that Matlab's `fmincon` implementation of the active set algorithm, the algorithm Gafarov, Meier, and Montiel Olea (2018) also use, solves the problem efficiently in

the single-shock case here.[14] For high-dimensional problems, we can run the algorithm in parallel, given independent posterior draws for $\mathbf{B}^{(d)}$ and $(\mathbf{\Sigma}^{\text{tr}})^{(d)}$. Unlike the literature, we quantify the importance of restrictions by computing Lagrange multipliers.

We can adapt the prior-robust algorithm to compute bounds for any well-defined function of the reduced form parameters and rotation vectors. Examples are implied policy rules (see Arias, Caldara, and Rubio-Ramírez (2019)), fiscal multipliers, or the FEV decomposition. Besides impulse responses, we focus here on the FEV decomposition. In Appendix C, we follow Uhlig (2003) to show that the conditional FEV for variable i at horizon H associated with the orthonormal vector $\mathbf{q}$ can be expressed as $\mathbf{q}'\mathbf{V}_{i,H}\mathbf{q}$ where $\mathbf{V}_{i,H} \equiv \sum_{h=0}^{H}(\mathbf{e}_i(\mathbf{B}_x^{(d)})^h(\mathbf{\Sigma}^{\text{tr}})^{(d)})'(\mathbf{e}_i(\mathbf{B}_x^{(d)})^h(\mathbf{\Sigma}^{\text{tr}})^{(d)})$. We can then compute bounds on the FEV contribution of any variable i up to horizon H by replacing the objective $\mathbf{e}'_j(\mathbf{B}_x^{(d)})^h(\mathbf{\Sigma}^{\text{tr}})^{(d)}\mathbf{q}$ in the previous algorithm with $\mathbf{q}'\mathbf{V}_{i,H}\mathbf{q}$. This approach is the algorithm used in Faust (1998) to assess whether the finding that monetary policy shocks only explain a small proportion of output is robust.

To summarize the prior-robust posterior, we report percentiles of the distribution of bounds. These percentiles yield equal-tailed credible regions, which may be conservative in their coverage, as Kline and Tamer (2016) discuss. In practice, when we compared these equal-tailed credible sets to the shortest credible sets proposed by Giacomini and Kitagawa (2018), we found only minor differences; see Figure D.2 in Appendix D. Since Bayesian practitioners commonly use equal-tailed credible sets, we therefore follow this simpler approach below.

3.5 *Fully Bayesian inference*

If a researcher has beliefs that provide information in addition to the sign restrictions, she might want to impose these beliefs. Here, we provide a framework for conducting inference under the belief that the rotation vector $\mathbf{q}$ is distributed uniformly over the unit n-sphere, conditional on lying in the identified set $\mathcal{Q}$. Fittingly, Uhlig (2017) labeled this belief "conditionally uniform."[15]

The conditionally uniform prior yields the following complete Bayesian model:[16]

$$p(\mathbf{y}^T, \mathbf{B}, \mathbf{\Sigma}, \mathbf{q}; \mathbb{L}) = \ell(\mathbf{B}, \mathbf{\Sigma}|\mathbf{y}^T)\pi_{\mathbf{B},\mathbf{\Sigma}}(\mathbf{B}, \mathbf{\Sigma})\pi_{\mathbf{q}}(\mathbf{q}|\mathbf{B}, \mathbf{\Sigma}; \mathbb{L}), \tag{3.6a}$$

---

[14]We experimented with different algorithms and solvers to ensure robustness of the results. We use `Knitro` 11.0 (Byrd, Nocedal, and Waltz (2006)) for higher dimensional cases to speed up computations. `Knitro` and `fmincon` offer the same algorithm. In our testing, both gave the same results, but `Knitro` cut computing times by two-thirds.

[15]Here, we strengthen the restrictions to involve strict inequalities, that is, we attempt to sample on the interior of the identified set only. Recall that we drop trivial rows in $\mathbf{W}$ with $\mathfrak{s} = 0$.

[16]Our prior implies that the marginal data density is unaffected by the prior over $\mathbf{q}$ when the identified set $\mathcal{Q}$ is never empty:

$$p(\mathbf{y}) = \int\int\int p(\mathbf{y}|\mathbf{B}, \mathbf{\Sigma})p(\mathbf{B}, \mathbf{\Sigma})p(\mathbf{q}|\mathbf{B}, \mathbf{\Sigma}; R(\cdot))\,d\mathbf{q}\,d\mathbf{B}\,d\mathbf{\Sigma}$$

$$= \int\int p(\mathbf{y}|\mathbf{B}, \mathbf{\Sigma})p(\mathbf{B}, \mathbf{\Sigma})\int p(\mathbf{q}|\mathbf{B}, \mathbf{\Sigma}; R(\cdot))\,d\mathbf{q}\,d\mathbf{B}\,d\mathbf{\Sigma}$$

$$= \int\int p(\mathbf{y}|\mathbf{B}, \mathbf{\Sigma})p(\mathbf{B}, \mathbf{\Sigma})\,d\mathbf{B}\,d\mathbf{\Sigma}.$$

$$\pi_{\mathbf{q}}(\mathbf{q}|\mathbf{B}, \boldsymbol{\Sigma}; \mathbb{L}) = \frac{\mathbb{1}\{W_f(\mathbf{B}, \boldsymbol{\Sigma}; \mathbb{L})\mathbf{q} \leq 0\}}{\displaystyle\int_{\mathcal{Q}} \mathbb{1}\{W_f(\mathbf{B}, \boldsymbol{\Sigma}; \mathbb{L})\tilde{\mathbf{q}} \leq 0\}\, d\tilde{\mathbf{q}}}. \tag{3.6b}$$

The conditionally uniform variant of Algorithm 1 generates draws from this prior.

While one can sample from the conditionally uniform prior using the accept-reject sampler in the literature (Uhlig (2005), Rubio-Ramírez, Waggoner, and Zha (2010)), this may require many proposals: To implement the conditionally uniform prior, we require a fixed number M of accepted draws of $\mathbf{Q}$ for each $(\mathbf{B}, \boldsymbol{\Sigma})$. Accepting M draws for $\mathbf{q}$ can require a large number of proposals when the identified set is small or the dimension of the VAR is large. In Appendix A.3, we therefore also describe a direct sampler, which we used for the computations here. Intuitively, the accept-reject sampler for a single shock draws vectors $\mathbf{z}$ from the rotation-invariant multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and accepts them if $\mathbf{Wz} < \mathbf{0}$. After acceptance, these draws are projected onto the unit sphere by rescaling to $\mathbf{q} = \frac{\mathbf{z}}{\|\mathbf{z}\|}$. Instead, our approach samples directly from the normal distribution $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ restricted to lie in the cone described by $\{\mathbf{z} : \mathbf{Wz} < \mathbf{0}\}$, before rescaling. While direct draws from this truncated distribution are typically not available, Li and Ghosh (2015) provided a simple Gibbs sampler.[17]

4. Application: Productivity news shocks

4.1 *Productivity news and industry data: Heterogeneity restrictions*

We show here how adding and exploiting heterogeneity yields credible sets that are substantially narrower than with standard restrictions. To incorporate heterogeneity restrictions, we include and rank the exposure of different industry returns to news shocks. The novel identifying assumption is that we restrict productivity news shocks to move the stock returns of the most innovative sectors more. Even though we use a model with more time series and, thus, more unknown parameters, our approach also yields narrower robust credible sets relative to a model with only an aggregate stock index. Inverting the order of the industries in the heterogeneity restrictions suggests, however, that our restrictions are not strong enough to distinguish TFP news from surprise shocks. Adding slope restrictions on TFP narrows the robust credible sets for key macro responses further.

4.1.1 *Data, specification, and identification*

Data.   Our benchmark data set consists of nine variables at quarterly frequency. We include four macro variables: Real gross value added in the business sector (output), utilization-adjusted TFP, hours worked in the business sector (all taken from Fernald (2014)) and consumer confidence. For the industry series, we use readily available real industry stock returns. To keep the estimation simple, we focus on the five-industry classification by Fama and French (1997), namely consumers, manufacturing, high tech,

---

[17]Geweke (1991) and Robert (1995) provided earlier, related Gibbs samplers. More efficient direct samplers such as Botev (2017), which uses a recursive sampler based on the LQ decomposition of the $\mathbf{W}$ matrix, are available when the number of restrictions is no larger than the dimension of $\mathbf{q}$.

(a) Relative to operating income   (b) Relative to assets
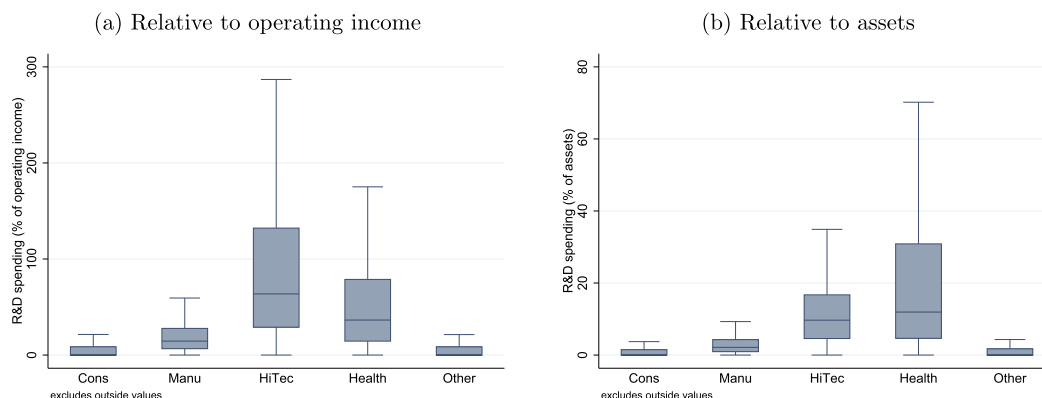

FIGURE 3. R&D intensity by industry in the five industry classification. The boxes show the median along with the interquartile range (IQR) of the R&D intensity for the five industries in the coarsest Fama and French (1997) classification. The upper (lower) whiskers end in the values just above the 75th (25th) percentile plus (minus) 1.5 times the IQR. We measure firm size either as the lagged 3-year moving average of operating income or total assets. Source: Compustat. U.S. firms, 1960–2015.

health, and other. For firms within each industry, we compute the distribution of R&D intensities, measured as the ratio of the three-year moving average of R&D expenses to a lagged measure of firm size. Figure 3 displays the distribution of the R&D intensity, pooled across firm-years, for each of the five industries using either gross operating income or total assets as a measure of size.[18] While we focus on five industries to keep the model parsimonious, we have also estimated and verified that our results hold using the finer ten industry classification.[19] Table 1 summarizes the benchmark data. The sample period is from 1960:Q1 to 2015:Q4.

Specification.   We use quarterly data in log-levels and allow for four lags. Our benchmark specification uses a flat prior. When we use a flat prior, we also include a deterministic quadratic trend, as recommended by Ramey (2016). Throughout, we take 500 reduced-form draws from the posterior. For each reduced-form draw, we generate 10,000 draws of the rotation vector **q** from the Gibbs sampler, with a thinning parameter of 10 that leaves us 1000 draws of **q**.

Identification.   We require a news shock to raise real GDP, hours, productivity, and consumer confidence as well as cumulative real stock returns. Based on the R&D intensities in Figure 3, we impose the following ranking restrictions on industry returns: (1) health and high tech returns increase more than those in manufacturing, and (2) manufacturing returns increase more than those in the consumer and other industries. We also impose that stock returns in the consumer and other industries increase. All restrictions are on impact and in the four subsequent quarters. Below we also report an extension

---

[18]We use Compustat data and drop observations with negative net sales, assets, or employment. We only keep U.S. firms. For our analysis, we winsorize the data at the 1st and 99th percentile year by year.

[19]Results are available upon request.

TABLE 1. Benchmark data and identifying restrictions in the VAR with industry-level data.

	Benchmark data	Sign restrictions	Heterogeneity restrictions
Macro (common)	Real output TFP Confidence Hours worked	Real output ≥ 0 TFP ≥ 0 Confidence ≥ 0 Hours worked ≥ 0	Real output ≥ 0 TFP ≥ 0 Confidence ≥ 0 Hours worked ≥ 0
Industry	FF-5 Consumers FF-5 Manufacturing FF-5 High Tech FF-5 Health FF-5 Other	FF-5 Consumers ≥ 0 FF-5 Manufacturing ≥ 0 FF-5 High Tech ≥ 0 FF-5 Health ≥ 0 FF-5 Other ≥ 0	FF-5 Manu $\geq$ FF-5 Consumers ≥ 0 FF-5 Manufacturing $\geq$ FF-5 Other ≥ 0 FF-5 High Tech $\geq$ FF-5 Manu FF-5 Health $\geq$ FF-5 Manu FF-5 Other ≥ 0

that imposes a (soft) zero restriction on initial TFP in the spirit of Beaudry and Portier (2006) and Barsky and Sims (2011).

Turning to the results, we discuss the impulse-responses first because this is where we impose the restrictions. We then turn to the FEV decomposition. Last, we analyze which restrictions are the most important and then conclude with robustness checks. The posterior plausibility of ranking restrictions is 99.8% without soft zero restrictions and 97% otherwise. Except for the extensions, we show results for heterogeneity and sign restrictions for the same reduced-form draws.

4.1.2 *Impulse response functions*   We now present results on impulse-response functions, (set-)identified with sign restrictions alone, and with added ranking restrictions. Our focus here is on the response of output, TFP, and hours.[20] The following figures present robust and fully Bayesian credible sets. Throughout, we contrast the traditional sign restrictions (SR) in the left panels for each response. The results with ranking restrictions are shown on the same scale on the right.

Heterogeneity restrictions (HR) narrow the robust credible set significantly. Figure 4 shows, in varying shades, the 98%, 95%, and 68% point-wise robust credible sets for the impulse-responses of output, TFP, and hours. Overlaid, in dashed lines, is the identified set at the posterior mean. HR enable us to make stronger statements about the output response than SR alone. Consider the 3-year horizon when the upper bound peaks with SR. The figure implies the output response is below 1.71% with 99% posterior probability, regardless about the prior over rotations one might hold. The associated width of the 98% robust credible set shrinks from 2.05 pp (= 1.71% − (−0.34%)) with SR to 1.42, a 30.7% reduction.

More generally, we find meaningful reductions in the width of the robust credible sets for all macro variables across all horizons. The reductions tend to increase at lower levels of the credible set. While these reductions can be deduced from Figure 4, Table D.1 in the Appendix D tabulates them for convenience. At the 2-year horizon, the reduction for output is 30.7% at the 98% level, but 39.2% at the 68% level. This compares to a reduction of 16.8% on impact and 31.9% after 6 years, at the 68% level. The patterns for

---

[20]The full set of results with additional specifications are available upon request.
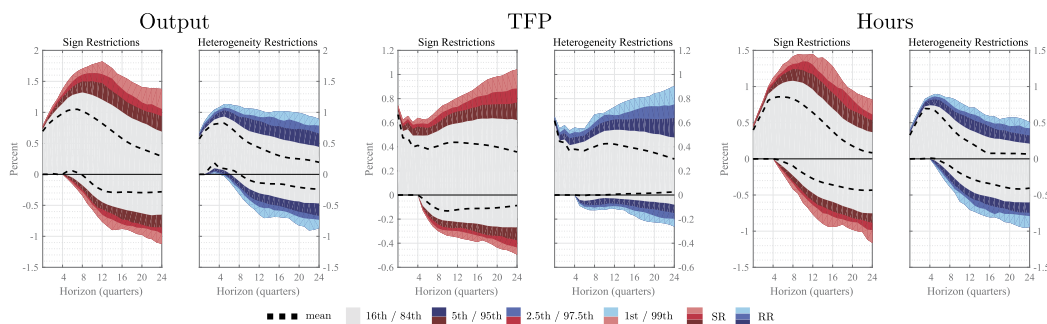
FIGURE 4. Prior-robust responses of output, TFP, and hours worked to a one standard deviation productivity news shock. With parameter uncertainty we can only bound the responses after impact. Heterogeneity restrictions sharpen the bounds: The peak upper bounds with heterogeneity restrictions are a third smaller for output and hours at the 99th percentile. Heterogeneity restrictions cut the lowest lower bound for TFP also by half.

hours and confidence are similar, while the largest reduction for TFP is near the 6-year horizon.

On the substantive side, the HR rule out economically significant reversals in TFP. While the robust credible set includes zero once the restrictions are lifted, with HR the posterior probability attached to negative response is much lower. Specifically, at the 6-year horizon, the posterior probability of a response of at least -0.11% is 90% with HR, but only 28% with SR. We can therefore rule out large declines following a positive TFP news shocks with HR, but not with SR.

HR also point to more modest responses of output and hours worked: The probability of increases in these measures of real activity of 1% or more at the 2-year horizon is less than 2.5% with HR. With SR, this probability could be as high as 28% for output and 72% for hours. Despite being smaller in size overall, heterogeneity restrictions also bound the output response away from zero at intermediate horizons: With HR, the posterior probability that output rises by at least 0.05% after 1 year is almost 98%. With SR, Figure 4 implies that this probability is less than $84\% - 56\%$, to be specific.

What are the implications of sign or heterogeneity restrictions under conditionally uniform priors? Figure 5 shows that a conditionally uniform prior over **q** sharpens inference substantially: The pointwise 98% credible sets all exhibit well-defined shapes. While we can say little about the shape of the TFP response while being robust to any prior, our fully Bayesian posterior implies that TFP increases in a hump-shaped fashion in response to a productivity news shock, plausibly reflecting technology diffusion. Inference about the hump is much sharper with HR. They imply a peak increase in TFP of 0.1% to 0.5% about 3 years after the initial shock with 95% probability.[21] This causes

---

[21]For comparison with the applied literature and because we also treat prior-robust identification, we present pointwise credible sets. See Inoue and Kilian (2013) and Montiel Olea and Plagborg-Møller (2019) for references on joint confidence sets. For output in Figure 5, we compute that for 98.6% of all draws with HR and 82.2% of all draws with SR the responses exhibit a hump, that is, they first rise and then fall in the first 8 quarters.
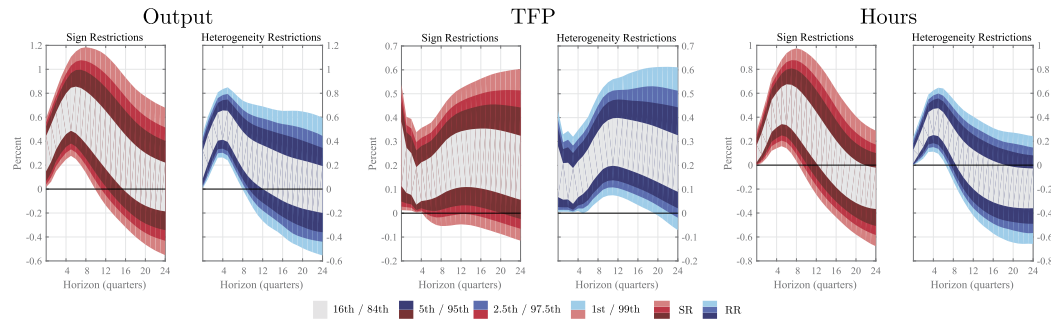
FIGURE 5. Fully Bayesian responses of output, TFP, and hours worked to a one standard deviation productivity news shock. Our fully Bayesian posterior implies that TFP increases in a hump-shaped fashion after a productivity news shock. The hump shape in TFP is well-defined only with heterogeneity restrictions. Heterogeneity restrictions also visibly reduce the credible sets for hours and TFP.

a hump-shaped expansion in output, peaking one year out between 0.3–0.8% with 95% probability, according to the model with HR or 0.3–1.1% with SR only. Hours worked peak at 0.15% to 0.6% (0.15% to 0.9% with SR) with 95% probability and then may turn negative: In the long term, the negative wealth effect on labor supply seems to offset the positive effect of productivity growth, returning output to trend. Overall, we see economically sensitive responses that are much sharper with heterogeneity restrictions.[22]

We now turn to the industry-responses of cumulative industry stock returns in Figure 6. HR on these responses yield tighter bounds on the macro variables. Mechanically, they also yield more precise inference about the industry responses themselves. The initial returns erode more quickly under heterogeneity restrictions in the other industries and the manufacturing industry than sign restrictions would indicate. We find similar patterns for all five industries and show here one of each category: the low R&D intensive other industries, manufacturing, and high tech. HR reduce the width of the 98th percent robust credible sets of all industry responses between 10% and 34% on impact and between 26% and 63% 1 year out (not shown).

4.1.3 *Forecast error variance decomposition*   The literature on news shocks often highlights that news shocks explain a significant part of the FEV in macro variables (e.g., Beaudry and Portier (2014)). Barsky and Sims (2011), and Kurmann and Sims (2017) advocate identifying productivity news shocks by maximizing the FEV. Here, we decompose the FEV using both prior-robust and conditionally uniform beliefs. The substantive conclusion that emerges is that, under some beliefs, news shocks could indeed drive

---

[22]The credible sets alone could obscure irregular posterior distributions, but Figure D.3 in the Appendix D shows that they do not. The posterior densities are unimodal and largely symmetric. The plot also confirms that the densities assign positive measure to almost the extremes of the distribution over identified sets, shown as thick lines underneath the zero line. On the substantive side, the densities show that for output and hours worked, the posterior mass shifts toward zero using heterogeneity restrictions. In contrast, the mass shifts toward positive values for TFP. For all variables, the densities are more concentrated with heterogeneity restrictions.
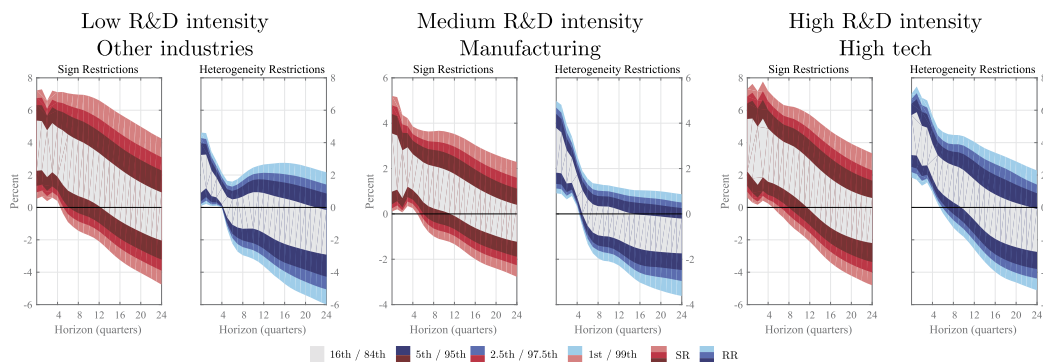
FIGURE 6. Fully Bayesian responses of (cumulative) industry returns to a one standard deviation productivity news shock. Heterogeneity restrictions rank the responses of stock returns of industries from zero to 4 quarters according to their R&D intensity to sharpen inference about macro variables. Heterogeneity restrictions mechanically imply larger response in more R&D intensive industries. These differences are absent with only sign restrictions.

most of the FEV in macro variables over some horizons. But the conditionally uniform belief points to a significant, yet more modest, role.

We first consider the identified set for the FEV contribution of productivity news to the macro shocks. For every reduced form parameter, there is always some prior over rotation vectors that yields a near-zero variance contribution. Figure 7(a) therefore only shows the posterior median upper bound for the contribution of productivity news, at horizons of up to 6 years. Given the lower bound of zero, this is also an estimate of the identified set. The upper line, the posterior median upper bound with sign restrictions, shows that SR alone are uninformative because the FEV contribution could be as high as 60% to 100% of the total FEV with 50% posterior probability, depending on the horizon. HR, in contrast, consistently lower the median upper bound for the FEV contribution to below 70%. At the 1-year horizon, the median reduction in upper bound for the FEV contribution is 32 pp for output. The 68% posterior credible set for the reduction in the upper bound, the shaded area, ranges from 20 pp to 38 pp.

While our prior-robust results show that news shocks can play a very large role even under HR, they play a modest role under conditionally uniform beliefs. Figure 7(b) shows the variance decomposition in this case. News shocks are still important, but far less so than under prior-robust beliefs: For output, news explains 14–28% of the total FEV at the one-year peak, and 11–27% of TFP after 6 years with HR. Compared with SR, HR suggest that news shocks are less important for output, hours, and confidence, but more important for TFP. This suggests that HR identify a productivity news shocks more precisely.

4.1.4 *Important restrictions*   We impose tighter restrictions to achieve sharper identification. Which of these restrictions matter? We use the Langrange multipliers from the prior-robust Algorithm 1(a) to quantify the importance of individual restrictions. This makes our results even more transparent.

(a) Prior-robust: Upper bound on FEV contribution of news shocks

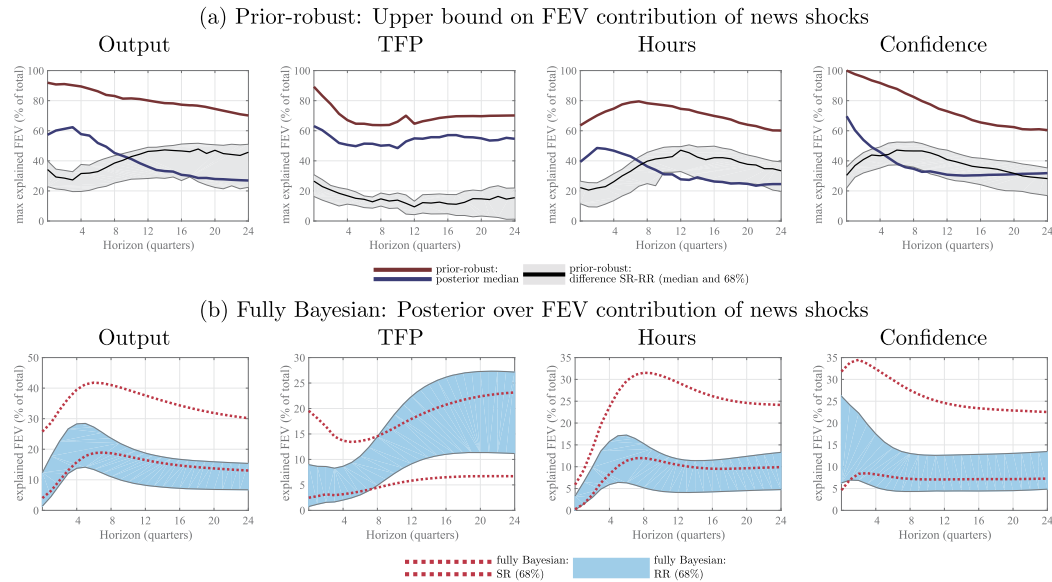(b) Fully Bayesian: Posterior over FEV contribution of news shocks

FIGURE 7. Forecast error variance contribution of productivity news shock: macro variables. FEV contributions are relative to the total FEV. Heterogeneity restrictions reduce the maximum role of news shocks (upper panel): With sign restrictions alone, news could explain all of the initial output and confidence FEV. Heterogeneity restrictions shrink the maximum FEV by 20–50 pp for output with 68% posterior probability. The reduction in the bound for TFP is up to 30 pp. The bounds remain wide in the short-run. With conditionally uniform beliefs (lower panel), the importance of news peaks at below 30% for output after 1 year with 68% probability, compared to the prior-robust result that with 50% posterior probability the response could be as high as 60%.

Lagrange multipliers quantify the role of restrictions in sharpening our inference by answering the question: How would impulse-responses change if we tightened a given restriction by a small amount? For example, how would the output response 2 years after a news shock change if we required consumer confidence not only to be positive initially, but bigger than ϵ? Or what if manufacturing stock returns had to increase more than consumer goods stock returns plus ϵ? Figure 8 answers these questions for all restrictions. The left panel shows the multipliers characterizing the lower bound, and the right panel shows those for the upper bound. Because multipliers depend on uncertain parameters, we show their distribution across reduced-form parameters. To simplify, we sum multipliers on restrictions across all horizons for which we impose them.

Heterogeneity restrictions are the most important for tightening the upper bound on the output response, but matter little for the lower bound, according to Figure 8. For example, the upper bound would not decrease if we imposed that consumer confidence had to rise more initially. In contrast, tightening the heterogeneity restriction that stock returns in the manufacturing industry rise relative to returns in the other industry would lower the upper bound of output by 0.11 to 0.23 pp. The restriction that manufacturing stocks rise more than consumer industry stocks still would lower the upper bound by
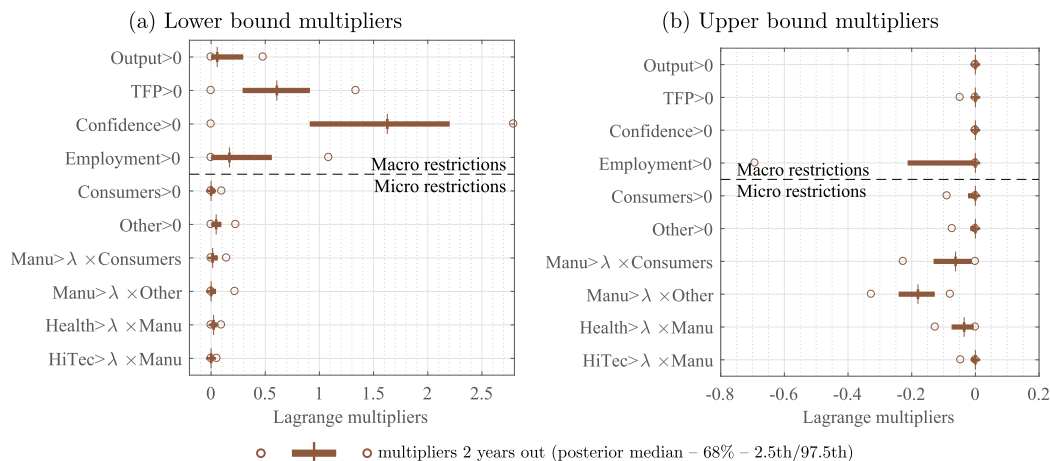
FIGURE 8. Lagrange multipliers on restrictions for output responses to news shocks after 2 years. This figure quantifies the importance of all sign and heterogeneity restrictions for the lower and upper bound of the output response at the 2-year horizon. It shows the distribution of Lagrange multipliers on all restrictions across reduced-form draws. The multipliers are summed across the restriction horizons $0, \ldots, \bar{H}$. Multipliers on upper bounds are negative as tighter restrictions lower the bound. For pinning down the upper bound, the heterogeneity restrictions on stock return industry data matter more than the macro sign restrictions. Restrictions on manufacturing are particularly important. In contrast, for the lower bound, sign restrictions on macro variables dominate.

0.02 to 0.14 pp. When we look at the determinants of the lower bound, the opposite picture emerges: The initial restrictions on macro variables, particularly consumer confidence, matter the most, while the industry-level restrictions hardly matter.

Our analysis of multipliers[23] also shows that standard SR on stock returns have little effect: The distribution (across reduced form draws) of differences in multipliers with $\lambda = 0$ and $\lambda = 1$ shows that multipliers on industry restrictions would be close to zero if we set $\lambda = 0$ and worked with pure sign restrictions. In contrast, multipliers on macro restrictions do not change systematically, highlighting the importance of HR.[24]

4.1.5 *Robustness*

Information set. Although we need more observables to implement HR, they still sharpen inference. More observables increase both the information set and the number of unknown parameters in the VAR. Implementing the larger VAR with sectoral stock returns and heterogeneity restrictions we find it to sharpen inference relative to a smaller VAR, despite the extra parameter uncertainty. The VAR has the same four macro variables, but only uses the mean of the sectoral returns (results hardly change with the Wilshire 5000). For output, hours, and confidence, the estimated IRFs are sharper with

---

[23]results available upon request

[24]We find, that the first and the last restriction horizon matter the most. Intuitively, when responses are monotone over the restriction horizon the first and last restriction determine the shape in between.

HR and industry-level data. For TFP, the credible sets are comparable at the 6-year horizon and sharper at short horizons.

Soft zero restrictions.   Beaudry and Portier (2006) and Barsky and Sims (2011) imposed the restriction that news cannot raise TFP immediately to identify news shocks. Here, we incorporate this assumption as a "soft" zero restriction on the initial TFP response.[25] We find that this extra restriction yields an additional set reduction: For output, this reduces the maximal FEV by an additional 10 pp to 15 pp compared with heterogeneity restrictions alone. The reduction for employment is 5 pp to 10 pp, while consumer confidence is hardly affected. By construction, the FEV for TFP that can be explained drops dramatically at short horizons but rises with the forecast horizon. The impulse-responses change little, except for TFP.

Other.   We have implemented four more checks: (1) more industries, (2) an informative reduced form prior, (3) a shorter restriction horizon, and (4) identifying both TFP news and surprise shocks.[26]

4.1.6 *Strength of ranking restrictions: Effects and inference*   Ranking restrictions give researchers the opportunity to impose both qualitative and quantitative restrictions: Response i is restricted to be stronger than λ times the response of j. In contrast, standard SR only require responses to be positive or negative, so that there is no role for an intensity. Throughout most of this paper, we take the stand that in most cases, qualitative ranking restrictions are the natural benchmark. $\lambda = 1$ reflects this conservative approach. In some cases, such as elasticity bounds, λ can be calibrated to implement these bounds. Researchers could, however, also let the data speak to the plausible magnitude of λ, as we now illustrate.

The data can inform the choice of λ through the plausibility of the implied restrictions. This criterion is conservative, since it is satisfied if there is any belief that a researcher could hold under which the identified set has positive measure. While conservative, we show that in our application the data are informative about the strength of the HR. Intuitively, as λ increases, the restrictions become stronger. When they are too strong, the associated identified set is empty for a large region of the parameter space. In practice, the plausibility starts to fall below unity once λ rises above one; see Figure D.4 in the Appendix.

While λs above unity are less plausible, they provide sharper identification in our application. Figure 9 illustrates that. The figure shows the output response under SR and HR, where $\lambda \in \{\frac{1}{3}, 1, 2\}$. Under each panel, we report the plausibility of each model. (See D.4 for $\lambda \in [\frac{1}{3}, 3]$.) The plausibility is, essentially, unity, except for $\lambda = 2$, when it is 29.2%. The loss in plausibility is associated with a reduction of the width of the robust credible set. At the 1-year and 2-year horizons, the 98% credible set shrinks by about 30% relative to our benchmark case of $\lambda = 1$. HR with $\lambda = \frac{1}{3}$ yield results similar to SR.[27]

---

[25]Formally, we impose on impact that output $> 10 \times$ TFP, in addition to TFP > 0.

[26]Results not reported here and available upon request.

[27]While we chose to report results separately for each λ for transparency, researchers could also integrate over λ if they have specific priors. Specifically, one could use prior weights and the plausibility in the data to form a (robust) posterior over the IRF that reflects uncertainty about the intensity of the restrictions.

Impulse response functions of output across varying intensity parameter λ

$\Pr_{SR} = 100\%$ $\Pr_{HR(\lambda)} = 100\%$ $\Pr_{HR(\lambda)} = 99.8\%$ $\Pr_{HR(\lambda)} = 29.2\%$
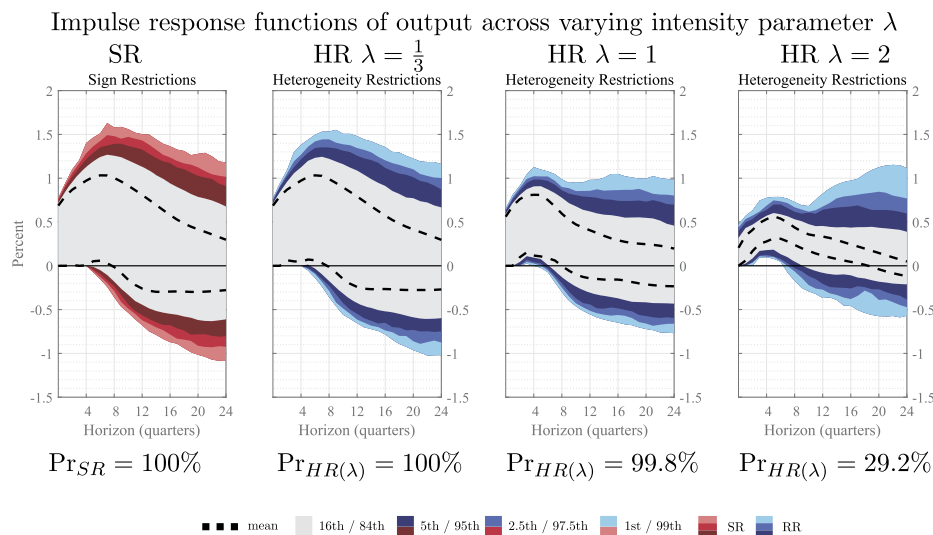
FIGURE 9. Prior-robust responses of output to a one standard deviation productivity news shock across varying values for intensity parameter $\lambda = \{0, \frac{1}{3}, 1, 2\}$. Intensities of λ above the benchmark value of $\lambda = 1$ yield to an increasingly smaller posterior plausibility $\Pr_{HR(\lambda)}$, but also narrower prior-robust credible sets. $\lambda = 1$ yields narrower credible sets than $\lambda = \frac{1}{3}$ with essentially the same posterior plausibility.

4.1.7 *Misspecification of ranking restrictions*   Misspecification of shocks can take two forms: (1) Wrong or overly tight restrictions can result in empty identified sets. Kline and Tamer (2016) and Giacomini and Kitagawa (2018) suggested the posterior probability of empty identified sets as a measure of misspecification. (2) Weak restrictions allow the identified to include a significant measure of linear combinations of other shocks. For example, a combination of expansionary demand and supply shocks can masquerade as monetary policy shocks if the restrictions are weak (Wolf (2020)).

Ranking restrictions allow us to go beyond the conservative test for empty identified sets. Of course, testing the posterior plausibility of the restrictions is a necessary first step. In addition, we can revert the rankings, in the spirit of placebo tests. If the restrictions imposed on the original model are strong, we expect the model with inverted restrictions to be rejected by the data. If not, one would be concerned that the heterogeneity restrictions are not operating through the channel advocated by the researcher. Alternatively, the viability of the inverted restrictions can reveal that the original restrictions are weak, so that the misspecification test has too little power.

Reverting the HR in our full sample suggests that the baseline restrictions are, indeed, too weak: They fail to isolate TFP news shocks. Excluding the early years from our sample, or imposing slope restrictions on TFP, gives our proposed restrictions bite—and allows us to reject the model with inverted restrictions as misspecified.

Specifically, inverting heterogeneity restrictions—so that returns on the less innovative sectors are restricted to rise the most in response to the identified shocks—yields impulse-responses of TFP that are mean-reverting (see Figure D.5 in the Appendix).

TABLE 2. Posterior probability of nonempty identified sets with reversed rankings and added slope restrictions.

			No slope restrictions		Added slope restrictions	
Sample start	Restriction horizon	Only SR	Actual HR	Inverted HR	Actual HR	Inverted HR
1960	1	100%	100%	100%	99.8%	99.8%
1960	3	100%	100%	100%	98.4%	85.5%
1960	5	99.2%	99.8%	100%	84.7%	40.0%
1983	1	100%	100%	100%	>99.9%	98.3%
1983	3	100%	100%	98.8%	98.9%	34.1%
1983	5	100%	99.2%	71.2%	80.2%	0.6%

Note: The table displays the posterior plausibilities of three different models, each for the full sample and the post 1983 sample. The results based on 2500 draws. With added slope restrictions and in the late sub-sample, the data appear inconsistent with the inverted HR.

This suggests that the original restrictions cannot discriminate between TFP news and TFP surprises. A natural remedy is to strengthen the restrictions by requiring that TFP rises initially after a TFP news shock. With this restriction, the posterior probability of nonempty identified sets is only 40% with the inverted HR when imposed for 5 quarters. With the actual HR it is 84.7%, as Table 2 shows.

Focusing on the data from 1983Q1 on, as we do in our second application, yields similar conclusions. The Compustat data motivating the HR is initially sparse: In the 1960s, we only have R&D expenditures for 10% of firms. The coverage is improves drastically in the 1970s, but given the number of Compustat firms, the years from 1983 to 2000 still feature 50% more observations than we have for 1970 to 1983. In this subsample, the actual HR have a probability of nonempty identified sets of 99.2%, compared to 71.2% for the inverted restrictions. Applying the slope restrictions in this subsample results in a probability of nonempty identified sets of only 0.6% with inverted restrictions, but 80.2% with the actual HR. We conclude that during periods when the data speak to the HR, and with slope restrictions, the actual HR are supported by the data, while there is little support for the inverted restrictions that we consider misspecified.

Strengthening the HR with slope restrictions on TFP also sharpens inference on the other responses Figure 10 shows that, mechanically, TFP increases more with the slope restrictions: The posterior probability that TFP is positive after 10 quarters is now at least 97.5%, whereas this bound on the posterior probability is less than 84% in the baseline model. For output, which is not subject to additional restrictions, we also obtain a narrower robust credible set: The posterior probability that output rises by at least 0.1% 1 year out and is positive throughout the first 1.5 years is at least 95% with the slope restrictions, but only at least 84% with the baseline restrictions.

4.2 *Productivity news and macro data: Slope restrictions and multiple shocks*

We now examine how much slope restrictions help to disentangle productivity news from productivity surprises. To do so, we use a smaller VAR that controls for nominal factors and restricts multiple shocks simultaneously. On the substantive side, our focus
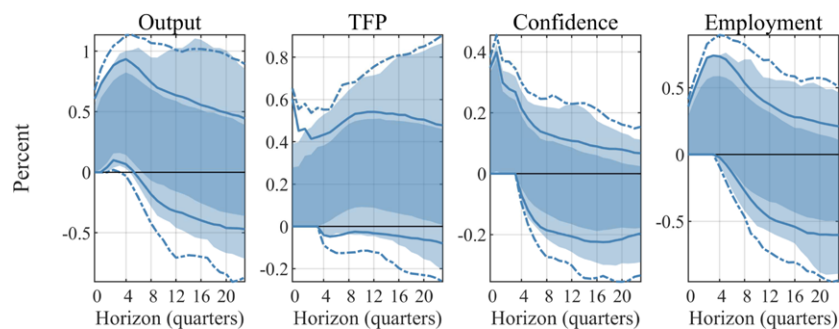
FIGURE 10. The effect of slope restrictions on TFP on the prior-robust credible sets of macro variables. The model contrasts the prior-robust 68% and 98% credible sets for the benchmark model with HR (shown in solid and dashed lines) with an extension that also includes slope restrictions on TFP (shown as shaded areas). Slope restrictions narrow the credible sets both for TFP and the other macro variables.

is on comparing our results for the impulse-responses and the characterization of the variance decomposition to the larger VAR. On the methodological side, this analysis illustrates our approach in the presence of multiple shocks.

4.2.1 Data, specification, and identification

Data.   Here, our data set consists of five macro variables at quarterly frequency: (1) real gross value added in the business sector (output), (2) utilization-adjusted TFP, (3) the real S&P500, (4) a measure of nominal interest rates, and (5) the CPI price level. The sample runs from 1983Q1 to 2016Q4. To accommodate the zero lower bound period, we use the 2-year treasury rate as a simple baseline measure, the shortest maturity used in Wright (2012). As a robustness check in the Appendix, we consider a measure of the shadow-rate of monetary policy (Wu and Xia (2016)).

Specification.   As before, we use quarterly data in log-levels, allow for four lags, include a deterministic quadratic trend, and use a flat prior in our benchmark specification. To test the identifying assumptions and as a robustness check, we also use a proper Minnesota prior. Throughout, we take 500 reduced-form draws from the posterior. For each reduced-form draw, we generate 50,000 draws of the rotation vector **q** from the Gibbs sampler, with a thinning parameter of 10.

Identification.   We identify a productivity news shock, a productivity surprise shock, and a monetary policy shock. As before, we impose the standard sign restrictions that TFP news increases output, TFP itself, and stock market valuations. TFP surprises also raise output and TFP, and lead to a drop in the price level. While the restrictions are distinct, a TFP surprise shock may alias as a news shock and vice versa, because the restrictions are mutually compatible. The slope restriction distinguishes the two shocks—by imposing that the news shock leads to a (weak) buildup in TFP, and the surprise shock to a mean-reversion. The slope restrictions are similar to the extension in the previous application. The monetary policy shock raises rates temporarily, and lowers output and

TABLE 3. Data and identifying restrictions in the 5-variable VAR.

Variable	Shock 1 "TFP news"	Shock 2 "TFP surprise"	Shock 3 "Monetary policy"
	Sign restrictions		
Output	$+ (0, \ldots, 4)$	$+ (0, \ldots, 4)$	$- (0, \ldots, 4)$
TFP	$+ (0, \ldots, 4)$	$+ (0, \ldots, 4)$	n/a
SP500	$+ (0, \ldots, 4)$	n/a	n/a
Nominal rate	n/a	n/a	$+ (0, \ldots, 2)$
CPI level	n/a	$- (0, \ldots, 4)$	$- (0, \ldots, 2)$
	Ranking restrictions		
TFP slope	$+ (0 \text{ vs } 1 \text{ and } 1 \text{ vs } 2)$	$- (0 \text{ vs } 1 \text{ and } 1 \text{ vs } 2)$	n/a
Nominal rate slope	n/a	n/a	$- (0 \text{ vs } 1 \text{ and } 1 \text{ vs } 2)$

Note: Restriction horizon in parenthesis.

the price level. For this shock, we impose the nominal restrictions for up to 2 quarters out and the output restrictions for up to 4 quarters out, to reflect the typically short half-life of monetary policy shocks (e.g., Smets and Wouters (2007)). The slope restriction also imposes that it be mean-reverting, unlike forward-guidance shocks (Campbell et al. (2012)). See Table 3.

The additional slope restrictions are overall consistent with the data, even though they are inconsistent with some regions of the reduced form parameter space. Specifically, we compute the posterior plausibility, that is, the probability that the identified set is nonempty when evaluated using the reduced-form posterior. With sign restrictions alone, the posterior plausibility is 100%. It drops slightly, to 90.4%, with the added slope restrictions given the flat prior used in most of our analysis; see the first row in Table 4. These results change little when we introduce proper priors. Specifically, we use a version of the Minnesota prior that is consistent with the trends that we also allow for. For the covariance matrix, our prior is a Wishart distribution with n degrees of freedom centered at the diagonal matrix with the standard deviations of the observed p observations prior to our sample start. For the coefficients, we use a sums-of-coefficients prior and a co-persistence prior with common hyperparameter ϕ.[28] With proper priors, the

TABLE 4. Posterior model probabilities and model comparison.

Prior	Posterior probability $\mathbb{E}_T^{\boldsymbol{\theta}}[\mathbb{1}\{\circ\}]$		Prior probability $\mathbb{E}_0^{\boldsymbol{\theta}}[\mathbb{1}\{\circ\}]$		2 ln Bayes factor
	SR	SR + RR	SR	SR + RR	SR + RR vs. only SR
Flat	100.0%	90.4%			
Weak ($\phi = \frac{1}{4}$)	100.0%	91.2%	80.2%	41.0%	1.16
Intermediate ($\phi = \frac{1}{2}$)	100.0%	94.4%	92.6%	73.0%	0.36
Stronger ($\phi = 1$)	100.0%	95.6%	98.6%	87.2%	0.16

Note: Estimated standard errors for prior and posterior probabilities are below 0.1% and below 0.03 for Bayes factors. For various priors over the reduced form parameters and a weak, proper prior over the covariance matrix, the data are consistent with the ranking restrictions.

---

[28] In the notation of Del Negro and Schorfheide (2011), we set the overall tightness hyperparameter $\lambda_1 = \phi$, $\lambda_2 = 2.5$ for the decay parameter to scale the prior standard deviations for lagged coefficients, $\lambda_3 = 1$

posterior plausibility of the model with added slope restrictions rises slightly, to up to 95.6%.

With proper priors, we can also use standard model comparison tools and evaluate the Bayes factor between the two models. Since the added slope restrictions are, a priori, less likely than the sign restrictions alone, we actually find that the data slightly favors the model with slope restrictions. Mechanically, the reason is that the posterior probability of the identified set rises more relative to the prior probability with slope restrictions than with sign restrictions alone. Overall, the added slope restrictions are thus consistent with the data.

4.2.2 *Impulse response functions*   Compared to pure sign restrictions, slope restrictions sharpen the inference, particularly at the posterior mean, as Figure 11 shows.[29] At the posterior mean, the identified set with slope restrictions is small enough that even its pointwise characterization implies a hump-shaped increase in output following a TFP news shock.[30] TFP builds up and remains above zero after 3 years. The identified set for the real SP500 also implies a hump-shaped rise, given the peak of its lower bound, which is narrowly above the upper bound on impact. The price index falls and returns to trend after 5 years. Sign restrictions alone reveal none of these patterns.

Parameter uncertainty is important, but slope restrictions are useful even in its presence. At the 2-year horizon, the robust credible sets shrinks between 11.8% and 33.6%, depending on the variable and the level of the credible set. While Figure 11 contains the
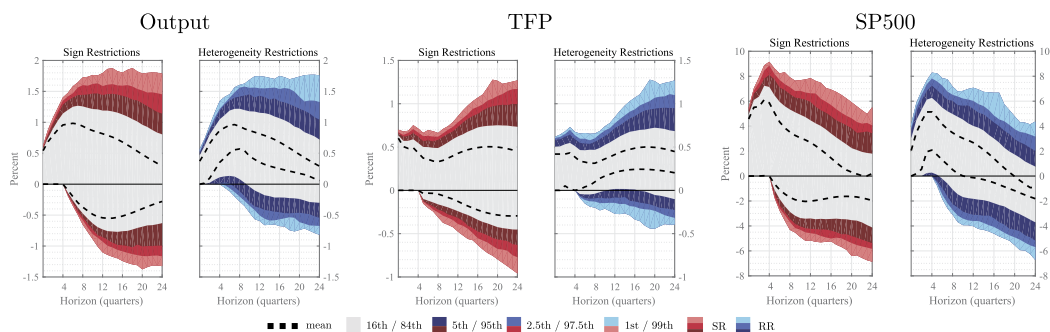

FIGURE 11. Prior-robust posterior over responses to a one standard deviation TFP news shock. Slope restrictions shrink the size of the identified set at the posterior mean significantly. For the three real variables, this is robust to parameter uncertainty, yielding narrower credible sets. The responses of output, TFP, and the SP500 are strictly positive with at least 84% posterior probability with slope restrictions even at unrestricted horizons.

denoting a priori number of observations used for obtaining the prior for the covariance matrix of error terms, $\lambda_4 = \phi$ denoting the hyperparameter for the sums-of-coefficients dummy observations prior and $\lambda_5 = \phi$ in denoting the hyperparameter for the copersistence dummy observations prior. See the summary in Table 4.

[29]Here, fewer reduced form draws underlie the inference on ranking restrictions, due to empty identified set $\mathcal{Q}$. In the Appendix, we use the same reduced-form draws for both and show that we still get sharper identified sets.

[30]The lower bound after 2 years is higher than the upper bound on impact and after 6 years.

underlying information for three of the five variables, Table D.2 reports the reductions in the credible set directly. Interestingly, even though the slope restrictions affect only TFP directly, the reductions are weakest for TFP itself at shorter horizons: At the 2-year horizon, the reductions range from 11.8% for TFP at the 98% level to 21.8% at the 68% level. At this horizon, they are strongest for output, whose robust credible set is reduced between 19.9% and 33.6%. This reduction in the credible set for the output response is similar in magnitude to what we found in the previous application.

Also similar to the previous application, the ranking restrictions rule out meaningful reversals in TFP following a news shock. With slope restrictions, the TFP response is positive 2 to 3 years after the news is revealed with at least 84% posterior probability. Without slope restrictions, the probability of a positive response could be as low as 5% over this horizon. This finding for TFP is unsurprising given that we restricted the TFP response to be increasing for 1 year. But the effects carry over to output: Sign restrictions alone do not allow us to sign the output response after the restrictions have ended. With slope restrictions, we find that output increases with about 90% posterior probability between 5 quarters and 8 quarters after the TFP news is revealed, irrespective of the beliefs over **q**.[31]

With the slope restrictions the fully Bayesian posterior credible sets resemble the identified set at the posterior mean in shape (Figure 12). By construction, the conditional uniform belief provides tighter credible sets than the prior-robust approach. In addition, it is easy to compute inference on the shape of the responses. For example, we find that both with sign restrictions alone and with added slope restrictions, the posterior probability of a hump-shaped output response is above 95%. Slope restrictions make the results significant over longer IRF horizons. For example, with slope restrictions the probability that output responds positively 3 years after the TFP news is 95%. With sign restrictions alone, this is true only at the 2-year horizon. Interestingly,
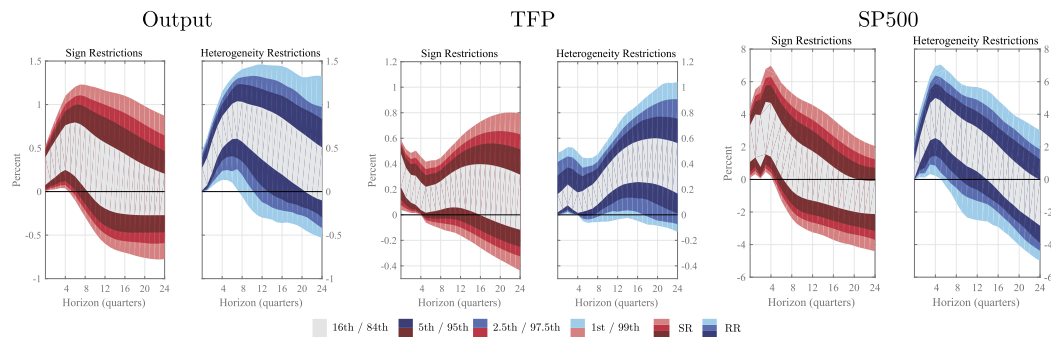

FIGURE 12. Fully Bayesian posterior over responses to a one standard deviation TFP news shock. Slope restrictions lead to longer-lasting and stronger increase in output, TFP, and the SP500 than pure sign restrictions.

---

[31]For the SP500, the restrictions have less bite. The probability that the SP500 responds strictly positively after 2 years is only at least 67% with slope restrictions. The 90% robust credible set still narrows by 17.1% at this horizon.

the credible sets for the output response are wider with the slope restrictions, reflecting an increase in the inferred magnitude of the output response, in line with the inferred stronger buildup in TFP. We conclude that slope restrictions as a form of ranking restrictions are useful for refining identification schemes.[32]

4.2.3 *Forecast error variance decomposition*   The set reduction for impulse-responses also translates to a set reduction for the upper bounds of the forecast error variance decomposition. Table 5 shows the contribution of the TFP news shock to the variance of output at various horizons.[33] With sign restrictions only, the posterior probability that TFP news account for 100% (up to rounding error) of the output variance on impact is at least 50%. With slope restrictions, this median upper bound falls to 32.1%. Note that the robust credible set always includes zero (not shown). Across draws, the posterior median reduction in the upper bound is 67.5 pp. The 68% credible set (across reduced form draws) for this reduction in bounds is 36.4% to 87.4%. After 1 year, the respective upper bounds are 92.2% and 58.6%, with a median reduction in bounds of 30.3%—similar to the reduction we found in the first application at this horizon. Here, however, the set reduction further diminishes at the 2-year horizon and largely disappears at the 6-year horizon. This contrasts with the slightly increasing set reductions in the first application.

The conditionally uniform Bayesian FEVD reflect both the narrower identified set as well as the stronger TFP build-up that we found in the IRF analysis. At short horizons, the 68% credible set is narrower with slope restrictions than with sign restrictions; see Table 5(b). At longer horizons, the bands are wider, and put more mass on bigger FEV contributions of productivity news shocks. While the 68th fully Bayesian

TABLE 5. Forecast error variance decomposition for output: TFP news shock.

	(a) prior robust upper bounds					(b) fully Bayesian 68% sets	
	SR	RR	Bound reduction			SR	RR
On impact	100.0	32.1	67.5	(36.4, 87.4)	On impact	(2.0, 52.6)	(0.1, 13.3)
4 qtrs out	92.2	58.6	30.3	(12.7, 59.4)	4 qtrs out	(6.4, 50.0)	(12.2, 47.3)
8 qtrs out	88.0	74.0	14.2	(2.7, 36.6)	8 qtrs out	(5.6, 47.6)	(25.9, 63.8)
24 qtrs out	76.9	70.0	3.6	(0.1, 18.6)	24 qtrs out	(6.4, 36.2)	(24.7, 62.8)

Note: Posterior medians (and 68% credible sets) of forecast error variance decompositions (FEVD), relative to total FEV at a given horizon. The set reduction is computed draw by draw, for draws accepted under ranking restrictions.

[32]The responses to the other two set-identified shocks are sensible: The posterior-robust credible sets remain too wide to identify any patterns beyond the identifying restrictions themselves. While we find that the slope restrictions also narrow the identified set for these shocks at the posterior mean, the gains in precision at the posterior mean are undone by the parameter uncertainty. This is unsurprising if mean-reverting TFP surprises and monetary policy surprises are the common shocks—the additional restrictions then add no qualitative information. For the fully Bayesian posterior for the other two shocks, we do see some quantitative gains in posterior precision for the output, TFP, and the CPI responses to a TFP surprise shock.)

[33]The results for TFP itself are very similar to those for output itself. For the other two shocks, the results with slope restrictions are very similar to those with sign restrictions only.

percentile lies below the median prior-robust upper bound, the 68th percentile is only about 10 pp. below the median upper bound with slope restrictions at longer horizons. These findings mirror what we found for the TFP variance decomposition using heterogeneity restrictions. Since ranking restrictions identify stronger productivity news shocks, the inferred variance contribution can be larger than with only sign restrictions.

4.2.4 *Important restrictions*   Figure 13 shows the Lagrange multipliers on the various restrictions for the lower bounds of the IRFs of output to TFP news 2 years out. For each response, we show the multipliers computed when the TFP news shock is identified separately and when it is identified jointly with two other shocks. We also find similar patterns for the upper bounds. These figures imply, for example, that an ε of tightening the slope restriction that TFP rises less on impact than after 1 quarter would raise the output response 2 years out by about 3.5ε at the posterior median. The 68% posterior excludes zero.

These plots imply that, in this application, restrictions on other shocks matter little for the identification of the TFP news shock: First, the multipliers on the restrictions that identify the TFP surprise shock itself are virtually the same whether we identify the shocks separately or jointly. Second, the multipliers on the restrictions on the other two shocks are close to zero with at least 68% posterior probability. Correspondingly, identifying the news shock jointly or separately makes little difference for the impulse-response functions.
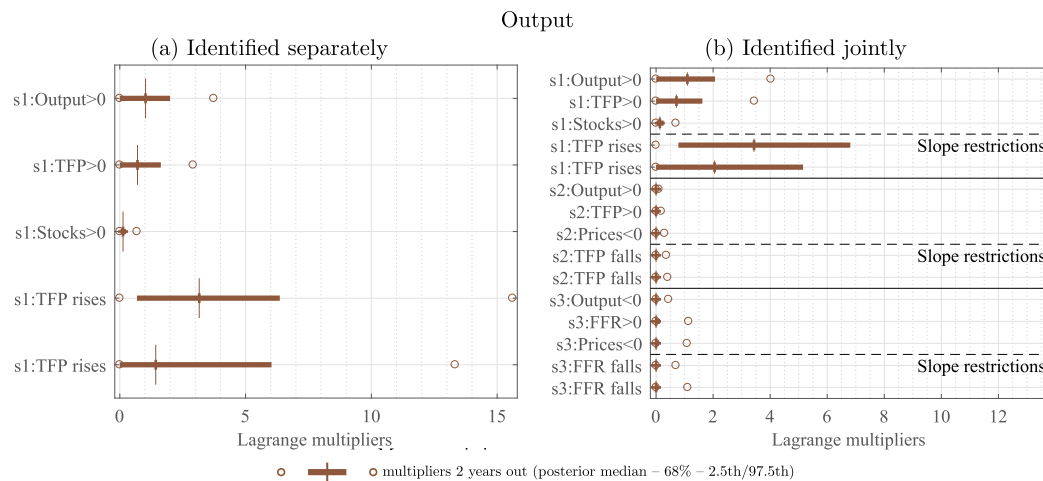

FIGURE 13. Lagrange multipliers for the lower bound of the 2-year output response, comparing joint and separate identification. "s1" refers to TFP news shocks, "s2" to TFP surprises, and "s3" to monetary policy shock restrictions. The multipliers for standard sign restrictions are summed across restriction horizons. Slope restriction have the largest multipliers. Multipliers on the restrictions on other shocks are zero for almost all reduced form draws, so that the identified set is pinned down from the restrictions on the productivity news shock itself.

5. Conclusion

A growing literature uses sign restrictions on impulse-response functions to set-identify shocks in VARs. We argue that in many applications, researchers have beliefs over additional qualitative relationships between impulse-response functions that can be used to sharpen inference. Here, we consider one class of such beliefs, which we refer to as ranking restrictions.

There are three types of ranking restrictions. First, ranking restrictions based on micro data, which we label heterogeneity restrictions. Formally, these are equivalent to elasticity bounds on macro variables, the second type. Both heterogeneity and elasticity restrictions impose a ranking on the relative magnitude of impulse-responses of different variables at the same horizon. Third, we introduce slope restrictions. They rank the responses of the same variables over different horizons. In small VARs, we show which features of the reduced form parameters lead to tighter identification compared to standard sign restrictions.

Whether sign or ranking restrictions narrow the identified set depends on the relationship between the restricted covariances between the forecast errors of variables conditional on the restricted shock, and its unconditional counterpart. The effectiveness of ranking restrictions is, therefore, an empirical manner. We illustrate their empirical relevance in two applications that analyze the effect of productivity news shocks on the economy. In both applications, ranking restrictions sharpen inference about macro variables such as output compared to just sign restrictions. This is not mechanical, since we do not impose additional restrictions on output.

Our results are based on Bayesian inference. We document their usefulness both in a prior-robust environment, which does not require us to specify beliefs over the distribution of the set-identified rotation vectors, and under fully specified Bayesian beliefs. However, there is a growing frequentist literature on inference in set-identified VARs and the restrictions we propose can be implemented equally well in the frequentist paradigm.

References

Amir-Ahmadi, P., and T. Drautzburg (2021), "Supplement to 'Identification and inference with ranking restrictions'." *Quantitative Economics Supplemental Material*, 12, https://doi.org/10.3982/QE1277. [3, 4]

Amir-Ahmadi, P. and H. Uhlig (2015), "Sign restrictions in Bayesian FaVARs with an application to monetary policy shocks." Working Paper 21738, National Bureau of Economic Research. [6, 10]

Arias, J. E., D. Caldara, and J. Rubio-Ramírez (2019), "The systematic component of monetary policy in SVARs: An agnostic identification procedure." *Journal of Monetary Economics*, 101, 1–13. [18]

Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2018), "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications." *Econometrica*, 86, 685–720. [8, 16]

Barsky, R. B. and E. R. Sims (2011), "News shocks and business cycles." *Journal of Monetary Economics*, 58 (3), 273–289. [21, 23, 27]

Baumeister, C. and J. D. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83 (5), 1963–1999. [3, 8, 16]

Beaudry, P. and F. Portier (2006), "Stock prices, news, and economic fluctuations." *American Economic Review*, 96 (4), 1293–1307. [2, 8, 21, 27]

Beaudry, P. and F. Portier (2014), "News-driven business cycles: Insights and challenges." *Journal of Economic Literature*, 52 (4), 993–1074. [23]

Border, K. C. (2013), "Alternative linear inequalities." Discussion paper, Caltech Division of the Social Sciences. [14]

Botev, Z. I. (2017), "The normal law under linear restrictions: Simulation and estimation via minimax tilting." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79 (1), 125–148. [19]

Byrd, R. H., J. Nocedal, and R. Waltz (2006), "KNITRO: An integrated package for nonlinear optimization." In *Large-Scale Nonlinear Optimization* (G. di Pillo and M. Roma, eds.), 35–39, Springer. [18]

Campbell, J. R., C. L. Evans, J. D. Fisher, and A. Justiniano (2012), "Macroeconomic effects of federal reserve forward guidance." *Brookings Papers on Economic Activity*, 43 (1), 1–80. [8, 31]

Canay, I. A. and A. M. Shaikh (2016), "Practical and theoretical advances in inference for partially identified models." CeMMAP working papers CWP05/16, Centre for Microdata Methods and Practice, Institute for Fiscal Studies. [16]

Canova, F. and G. De Nicolo (2002), "Monetary disturbances matter for business fluctuations in the G-7." *Journal of Monetary Economics*, 49 (6), 1131–1159. [1]

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999), "Monetary policy shocks: What have we learned and to what end?" In *Handbook of Macroeconomics* (J. B. Taylor and M. Woodford, eds.), Handbook of Macroeconomics, Vol. 1, 65–148, Elsevier. Chapter 2. [8]

De Graeve, F. and A. Karas (2014), "Evaluating theories of bank runs with heterogeneity restrictions." *Journal of the European Economic Association*, 12 (4), 969–996. [2]

Del Negro, M. and F. Schorfheide (2011), "Bayesian macroeconometrics." In *The Oxford Handbook of Bayesian Econometrics* (J. Geweke, G. Koop, and H. van Dijk, eds.), 293–389, Oxford University Press. [31]

Fama, E. F. and K. R. French (1997), "Industry costs of equity." *Journal of Financial Economics*, 43 (2), 153–193. [2, 19, 20]

Faust, J. (1998), "The robustness of identified VAR conclusions about money." *Carnegie-Rochester Conference Series on Public Policy*, 49 (1), 207–244. [1, 3, 16, 17, 18]

Fernald, J. G. (2014), "A quarterly, utilization-adjusted series on total factor productivity." FRBSF Working Paper 2012-19 (updated March 2014), FRB SF. [19]

Gafarov, B., M. Meier, and J. Montiel Olea (2016), "Projection inference for set-identified SVARs." Discussion paper, New York University. [2]

Gafarov, B., M. Meier, and J. Montiel Olea (2018), "Delta-method inference for a class of set-identified SVARs." *Journal of Econometrics*, 203 (2), 316–327. [2, 17]

Geweke, J. (1991), "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities." In *Computing Science and Statistics: Proceedings of the 23RD Symposium on the Interface*, 571–578. [19]

Giacomini, R., T. Kitagawa, and A. Volpicella (2017), "Uncertain identification." Discussion paper, Cemmap Working Paper CWP18/17. [3, 14]

Giacomini, R. and T. Kitagawa (2018), "Robust Bayesian inference for set-identified models." Discussion paper, Centre for Microdata Methods and Practice (cemmap). [2, 3, 14, 15, 16, 17, 18, 28]

Granziera, E., H. Moon, and F. Schorfheide (2018), "Inference for VARs identified with sign restrictions." *Quantitative Economics*, 9 (3), 1087–1121. [2, 8, 9, 14, 16]

Inoue, A. and L. Kilian (2013), "Inference on impulse response functions in structural VAR models." *Journal of Econometrics*, 177 (1), 1–13. [22]

Kilian, L. and D. P. Murphy (2012), "Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models." *Journal of the European Economic Association*, 10 (5), 1166–1188. [2, 8]

Kilian, L. and X. Zhou (2018), "Structural interpretation of vector autoregressions with incomplete information: Revisiting the role of oil supply and demand shocks: Comment." CEPR Discussion Papers 13068, C.E.P.R. Discussion Papers. [3]

Kline, B. and E. Tamer (2016), "Bayesian inference in a class of partially identified models." *Quantitative Economics*, 7 (2), 329–366. [3, 14, 15, 18, 28]

Kurmann, A. and E. Sims (2017), "Revisions in utilization-adjusted TFP and robust identification of news shocks." NBER Working Papers 23142, National Bureau of Economic Research. [2, 23]

Li, Y. and S. K. Ghosh (2015), "Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints." *Journal of Statistical Theory and Practice*, 9 (4), 712–732. [19]

Montiel Olea, J. L. and M. Plagborg-Møller (2019), "Simultaneous confidence bands: Theory, implementation, and an application to SVARs." *Journal of Applied Econometrics*, 34 (1), 1–17. [22]

Piffer, M. (2016), "Assessing identifying restrictions in SVAR models." DIW Discussion Papers 1563, DIW Berlin. [14]

Plagborg-Møller, M. (2019), "Bayesian inference on structural impulse response functions." *Quantitative Economics*, 10 (1), 145–184. [2, 8]

Ramey, V. A. (2011), "Identifying government spending shocks: It's all in the timing." *The Quarterly Journal of Economics*, 126 (1), 1–50. [8]

Ramey, V. A. (2016), "Macroeconomic shocks and their propagation." NBER Working Papers 21978, National Bureau of Economic Research. [20]

Robert, C. (1995), "Simulation of truncated normal variables." *Statistics & Computing*, 5 (2), 121–125. [19]

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010), "Structural vector autoregressions: Theory of identification and algorithms for inference." *Review of Economic Studies*, 77 (2), 665–696. [3, 19]

Schorfheide, F. (2016), "Macroeconometrics—a discussion." Discussion. [17]

Sims, C. A. (1980), "Macroeconomics and reality." *Econometrica*, 48 (1), 1–48. [1]

Smets, F. and R. Wouters (2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97 (3), 586–606. [31]

Uhlig, H. (1994), "What macroeconomists should know about unit roots: A Bayesian perspective." *Econometric Theory*, 10 (3/4), 645–671. [12]

Uhlig, H. (2003), "What moves GNP?" Draft, Humboldt Universität zu Berlin. [18]

Uhlig, H. (2005), "What are the effects of monetary policy on output? Results from an agnostic identification procedure." *Journal of Monetary Economics*, 52 (2), 381–419. [1, 3, 5, 15, 19]

Uhlig, H. (2017), "Shocks, sign restrictions, and identification." In *Econometric Society Monographs*, Vol. 2, 95–127, Cambridge University Press. [2, 3, 8, 18]

Wolf, C. K. (2020), "SVAR (mis-)identification and the real effects of monetary policy shocks." *American Economic Journal: Macroeconomics*, 12 (4), 1–32. [28]

Wright, J. H. (2012), "What does monetary policy do to long-term interest rates at the zero lower bound?" *The Economic Journal*, 122 (564), F447–F466. [30]

Wu, J. and F. Xia (2016), "Measuring the macroeconomic impact of monetary policy at the zero lower bound." *Journal of Money, Credit and Banking*, 48 (2–3), 253–291. [30]