Comment

Mark W. Watson, *Princeton University and NBER*

In this paper, Debortoli, Galí, and Gambetti offer compelling empirical evidence that extraordinary actions taken by the Federal Reserve were able to shield the macroeconomy from many of the policy constraints associated with the zero lower bound (ZLB) on nominal interest rates. As Debortoli et al. argue, if these extraordinary actions had been ineffective, the United States would have witnessed a change in the volatility of macro aggregates and a change in their response to specific nonfinancial shocks. Yet volatility and impulse responses remained largely unchanged during the ZLB period.

There is no one more qualified than the paper's first discussant to discuss the ZLB, the Fed's actions, and their effects on the macroeconomy. With this in mind, I will offer no comments of substance about this excellent paper, beyond the observation that I am in agreement with Debortoli et al.'s overall empirical conclusions. Instead, I will focus my comments on a methodological issue: statistical inference in sign-restricted structural vector autoregressions (SVARs), which is one of the methods used in Debortoli et al.'s paper.

Sign-restricted SVARs are an increasingly popular method for estimating dynamic causal effects in macroeconomics. Many researchers use a variant of Uhlig's (2005) Bayes method for imposing these sign restrictions and conducting inference. This method has both strengths and weaknesses. The strengths are widely recognized by macroeconomists but the weaknesses far less so. This discussion explains and highlights these weaknesses.

I make two initial comments. First, Debortoli et al. use a sophisticated time-varying SVAR identified by both long-run equality restrictions and shorter-run sign restrictions. To keep things simple, I will focus on a time-invariant SVAR. Second, there is nothing original in my comments beyond a few numerical calculations. Sign-restricted SVARs are a special

978-0-226-70789-1/2019/2019-0303$10.00

case of "set identification," which is well studied in econometrics (Manski 2003 is a classic reference). The issues I highlight for Bayes inference in sign-restricted SVARs are derived and discussed in Baumeister and Hamilton (2015); many of my comments are simply a nontechnical summary of their analysis.

Why Sign-Restricted SVARs Are Different from Standard SVARs

I begin by introducing some familiar notation. Let Y_t denote the vector of observed data. (In the Debortoli et al. paper, $Y_t = (\Delta(y_t - n_t), n_t, \Delta p_t, i_t^{\text{Long}})'$, where (y, n, p) denote the logarithms of output, employment, and the price level, and i^{Long} is the 10-year Treasury bond rate, all for the United States.) The VAR for Y_t is

$$Y_t = A_0 + A(\mathrm{L})Y_{t-1} + u_t, \tag{1}$$

where A_0 is the intercept, $A(\mathrm{L})$ is the VAR lag polynomial, and u_t is the VAR's one-period-ahead forecast error ("innovation"), which is serially uncorrelated with covariance matrix Σ_u. The forecast errors u_t are linearly related to a vector structural shocks ε_t,

$$u_t = Q\varepsilon_t, \tag{2}$$

where Q is nonsingular. (In the Debortoli et al. paper, the structural shocks are $\varepsilon_t = (\varepsilon_t^{\text{Technology}}, \varepsilon_t^{\text{Demand}}, \varepsilon_t^{\text{Monetary policy}}, \varepsilon_t^{\text{Supply}})'$.) The distributed lag relating the observed data to the structural shocks is $Y_t = \tilde{A}_0 + (I - A(\mathrm{L}))^{-1}Q\varepsilon_t$, and the impulse responses, $\partial Y_{t+h}/\partial\varepsilon_t$ correspond to the various elements of the lag polynomial $(I - A(\mathrm{L}))^{-1}Q$.

A common parameterization of the SVAR uses standardized structural shocks, so the covariance matrix of ε_t is $\Sigma_\varepsilon = I$. With this "unit-standard deviation" parameterization, equation (2) can be written as

$$u_t = \Sigma_u^{1/2}R\varepsilon_t, \tag{3}$$

where $\Sigma_u^{1/2}$ is the Cholesky factor of Σ_u and R is a "rotation matrix" (i.e., a matrix with $RR' = I$). With this parameterization, the impulse responses are $(I - A(\mathrm{L}))^{-1}\Sigma_u^{1/2}R$. The values of $A(\mathrm{L})$ and $\Sigma_u^{1/2}$ can be estimated from the VAR in equation (1), leaving R to be determined by a researcher's a priori restrictions on the way the structural shocks affect the observed data.

In standard SVAR analysis, these a priori restrictions take the form of equality restrictions (e.g., $\partial Y_{i,t+h}/\partial\varepsilon_{j,t} = 0$ for some values of i, j, and h)

that uniquely determine R and the impulse responses. In contrast, sign-restricted SVARs rely on inequality restrictions (e.g., $\partial Y_{i,t+h}/\partial \varepsilon_{j,t} \geq 0$) that place set-valued constraints on R but do not uniquely determine its value. Each value of R in this "identified set" leads to different impulse responses, so the impulse responses are also set-identified. Some SVAR analyses, like the Debortoli et al. paper, use a combination of equality and sign restrictions that determine the value of some elements of R and place set-valued restrictions on other elements.

Some generic notation will streamline the discussion. Let μ denote the set of parameters that characterize the probability distribution of Y_t. In the VAR model, $\mu = (A, \Sigma_u)$ characterizes the data's Gaussian likelihood, where A denotes the parameters in A_0 and $A(\mathrm{L})$. Let θ denote the parameters of interest. In the SVAR, θ might denote a set of impulse responses $\partial Y_{i,t+h}/\partial \varepsilon_{j,t}$ for particular values of i, j, and h. In standard SVAR analysis, the identifying restrictions uniquely determine the value of θ from μ, that is $\theta = g(\mu)$ for some function g. In the jargon of econometrics, the value of θ is "point identified." In contrast, in sign-identified SVAR analysis, the value of θ is restricted to lie in a set that depends on μ, that is, $\theta \in G(\mu)$, where G is a set-valued function. In this case, θ is said to be set identified.

Bayes and Frequentist Inference for Set-Identified Parameters

Set identification presents challenges for statistical inference. In point-identified models, when the value of μ is known, then θ is known. Absent sampling uncertainty, Bayes and frequentist inference are trivially identical. And, when sampling uncertainty in μ is small and approximately normally distributed (as it is when the sample size is large), Bayes and frequentist inference often coincide when standard Bayes priors are used. This rationalizes large-sample Bayes analysis for frequentists and vice versa.

This large-sample Bayes-frequentist coincidence disappears in set-identified models. An example, much simpler than the SVAR, highlights the key issues. Suppose you have data on two variables $Y_t = (Y_{1,t}, Y_{2,t})'$ with

$$Y_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, I_2 \right) \tag{4}$$

for $t = 1, \ldots, T$, and with $\mu_1 < \mu_2$. In this example, the probability distribution for Y_t is completely determined by $\mu = (\mu_1\ \mu_2)'$. Suppose interest focuses on a parameter θ, and all that is known is that $\mu_1 \leq \theta \leq \mu_2$, so that θ is set-identified.

Figure 1 summarizes frequentist inference about θ. Figure 1*a* shows small-sample inference: the estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ are the sample means of the data; they contain sampling error, so the endpoints of the identified set for θ are uncertain. Figure 1*b* shows large-sample frequentist inference: sampling error disappears when T is large so the values of μ_1 and μ_2 and the identified set for θ are pinned down.

The corresponding figure for Bayes inference is shown in figure 2. Bayes inference requires a prior, say $f(\theta, \mu)$, and a likelihood, say $f(Y|\mu)$. The resulting posterior for θ is

$$f(\theta|Y) = \int f(\theta|\mu) f(\mu|Y) d\mu, \tag{5}$$

where $f(\theta|\mu)$ is the prior for θ conditional on μ and $f(\mu|Y)$ is the posterior for μ. Figure 2*a* shows the posterior $f(\theta|Y)$ for a hypothetical prior and data. As in equation (5), the posterior is the prior $f(\theta|\mu)$ averaged over the values of μ from the posterior. Figure 2*b* shows the large-sample posterior: here, as in the frequentist case, uncertainty about the value of μ has vanished, so the Bayes posterior coincides with the Bayes prior, that is, $f(\theta|Y) = f(\theta|\mu)$, where the prior is evaluated at the true value of μ.

For both frequentist and Bayes inference, the data are informative about the value of μ, perfectly so when $T \to \infty$. For frequentist inference, this yields a set of values for θ, $\mu_1 \leq \theta \leq \mu_2$, that are consistent with the data. One of the values in this identified set corresponds to the true value of θ, the other values do not, but nothing can be said about how likely one of these values is relative to another. In contrast, for Bayes inference, the prior $f(\theta|\mu)$ shows that some values of θ may be more likely than other

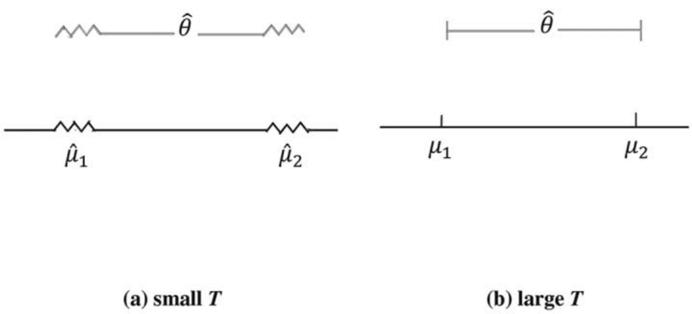

Fig. 1. Frequentist inference about θ. Jagged lines in *a* connote uncertainty about the values of μ_1 and μ_2.

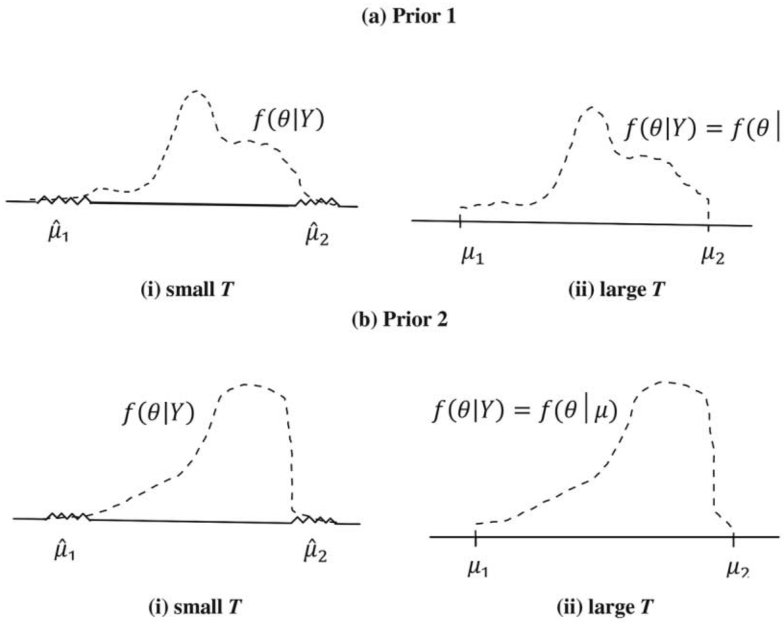

Fig. 2. Bayes inference about θ. (*a*) Prior 1. (*b*) Prior 2. Jagged lines above $\hat{\mu}_1$ and $\hat{\mu}_2$ in *a* and *c* connote uncertainty about the values of μ_1 and μ_2.

values. Importantly, this information comes solely from the prior—not the data—so different priors yield different inference, even when confronted with the same (possibly infinitely large) data set. This is highlighted in figure 2*b*, which uses the same hypothetical data and value of μ as figure 2*a* but uses a different prior $f(\theta|\mu)$. Bayes inference about the value of θ will be different using this new prior, even when confronted with an arbitrarily large sample of data (compare fig. 2*a.ii* and 2*b.ii*).

What are we to make of this simple example? Focusing on the large-sample results, I see three lessons. First, frequentist inference is informative when the identified set is small: if the data indicate that μ_1 is close to μ_2, then the restriction $\mu_1 \leq \theta \leq \mu_2$ says a lot about the value of θ. Second, Bayes inference is potentially more useful than frequentist inference because it informs us about the relative likelihood of values of θ within the identified set. For example, the median, mean, and 95% "credible" intervals for θ can be computed using Bayes methods but have no frequentist counterparts. Third, Bayes inference over the identified set is completely determined by the prior. Thus, conclusions drawn from Bayes analysis are persuasive only to the extent that the prior is credible. A corollary of this third lesson is that it is impossible to evaluate the conclusions from

Bayes analysis without knowing and evaluating the prior used in the analysis.

Identified Sets and Priors in a Version of the Debortoli et al. SVAR

To apply these lessons to the Debortoli et al. SVAR, I conducted a numerical exercise. In my experiment, I estimated a version of the authors' four-variable VAR using data from 1984 to 2008. This yielded estimates of the VAR parameters A_0, $A(L)$, and Σ_u; these are the μ-parameters for this exercise. Holding μ fixed, I then used the Debortoli et al. equality and sign restrictions to compute the identified sets for the impulse response functions (IRFs; θ for this exercise). I also used their prior to compute the posterior $f(\theta|Y)$. Because μ was held fixed in my exercise (i.e., I ignored sampling error in the VAR parameters), the posterior corresponds to the prior, that is, $f(\theta|Y) = f(\theta|\mu)$.

Figure 3 shows the identified sets for the IRFs for the three shocks identified by sign restrictions. The authors' identifying restrictions turn out to be (remarkably) informative about the values of these IRFs. For example, figure 4*a* shows the identified set for the response of output after 4 quarters (y_{t+4}) to a 1-standard negative demand shock ($\varepsilon_t^{\text{Demand}}$), that is, for the parameter $\theta = \partial y_{t+4}/\partial \varepsilon_t^{\text{Demand}}$. The identified set is $-0.27 \leq \theta \leq 0$. The upper bound of the set ($\theta \leq 0$) is one of the authors' sign restrictions, but the lower bound ($\theta \geq -0.27$) follows from the value of μ computed

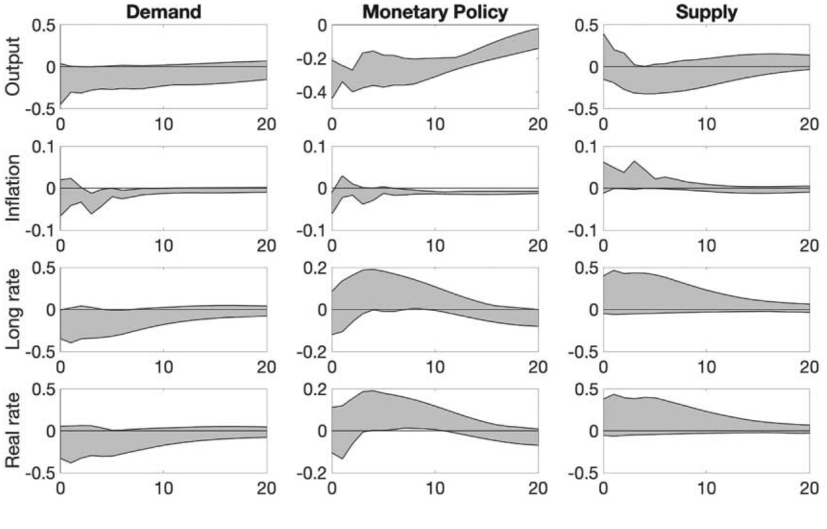

Fig. 3. Identified sets for the four-variable SVAR.

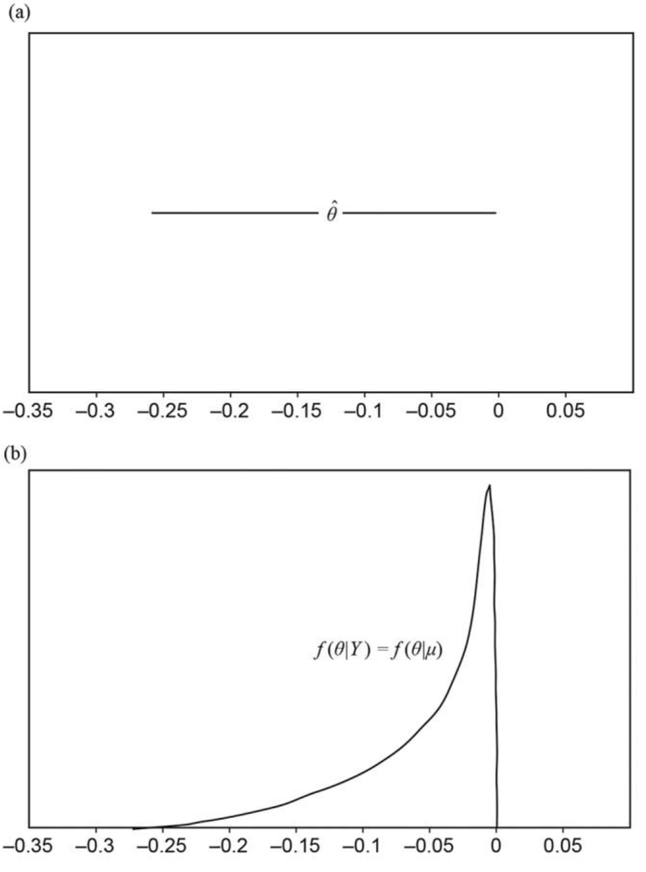

Fig. 4. Inference about $\theta = \partial \text{Output}_{t+4}/\partial \varepsilon_t^{\text{Demand}}$: (*a*) frequentist inference: the identified set; (*b*) Bayes inference: the truncated prior = posterior for θ.

from the data, together with the equality and sign restrictions. Looking across the panels in figure 3 leads me to conclude that the combination of sign and equality restrictions used by Debortoli et al. impose sufficient structure on the data to greatly narrow the range of IRFs. Frequentist inference is informative.

What more do we learn from Bayes inference? As in the large-sample version of the mean example in the last section, the posteriors for the impulse responses coincide with their priors over the identified sets. Thus, the question becomes what prior was used in the Debortoli et al. SVAR. In this paper, as in many sign-restricted VARs, the authors follow Uhlig (2005) and construct these priors by a flat (Haar) prior on the columns of the rotation matrix R truncated to satisfy the sign restrictions on θ. This prior for R induces a prior for the impulse responses, and this impulse response prior is easily computed using numerical methods.

Figure 4*b* shows the implied prior for $\theta = \partial y_{t+4}/\partial\varepsilon_t^{\text{Demand}}$ over the the identified set shown in figure 4*a*. Notice that the prior is quite informative about the effect of the demand shock on output. For example, the prior puts roughly 60% of its mass on $-0.05 \leq \theta \leq 0$ and less than 2% of its mass on values with $-0.27 \leq \theta \leq -0.20$. Comparing figure 4*a* and 4*b*, frequentist inference says that $-0.27 \leq \theta \leq 0$ and says nothing more. Bayes inference sharpens this, by saying that it is 30 times more likely that $-0.05 \leq \theta \leq 0$ than $-0.27 \leq \theta \leq -0.20$. Importantly, this latter conclusion follows from the authors' prior, not from any information in the data.

Figure 5 shows the implied priors for the other variables to the other shocks, again at the 4-quarter horizon. Here, too, the "flat" prior on R together with the authors' equality and sign restrictions imply quite informative priors on the impulse responses. For example, the prior implies that demand shocks have relatively small effects on output and prices but large effects on interest rates; in contrast, monetary policy shocks have relatively small effects on interest rates and inflation but large effects on output.

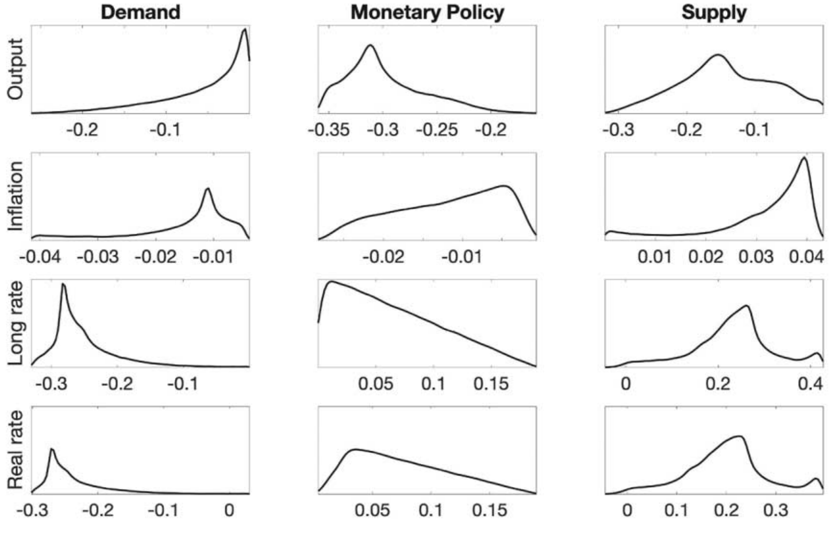

Fig. 5. Truncated priors (= posteriors) for $\partial Y_{t+4}/\partial\varepsilon_t$. The figure shows the prior on the four-period-ahead impulse responses induced by a flat prior on R and the equality and sign restrictions.

Are these priors reasonable? Maybe they are, and maybe they are not. In any event, the credibility of the conclusions reached using Bayes methods depends critically on the answer to this question. Given the key role played by the priors, it is remarkable that priors are not reported in the vast majority of papers using sign-restricted SVARs (including the Debortoli et al. paper). Simply put, it is impossible to evaluate the "point estimates" or "error bands" for IRFs without evaluating the priors.

To complete the numerical calculations, figure 6 shows the identified sets for the impulse responses previously shown in figure 3 together with selected percentiles of Bayes priors/posteriors. I have drawn the figures to highlight the point estimates (prior/posterior median), 68% error bands (prior/posterior 16% and 84% percentiles), and 95% bands. Readers are used to associating error bands with sampling uncertainty, but in large-sample sign-restricted SVARs, these error bands summarize the researchers' prior uncertainty, not sampling uncertainty.

Additional Comments and Recommendations for Practice

Let me end with two additional comments and some recommendations for practice. The first comment involves the unit standard deviation normalization for the structural shocks ($\Sigma_\varepsilon = I$) widely used in SVAR analysis (and in the Debortoli et al. paper). This normalization means that

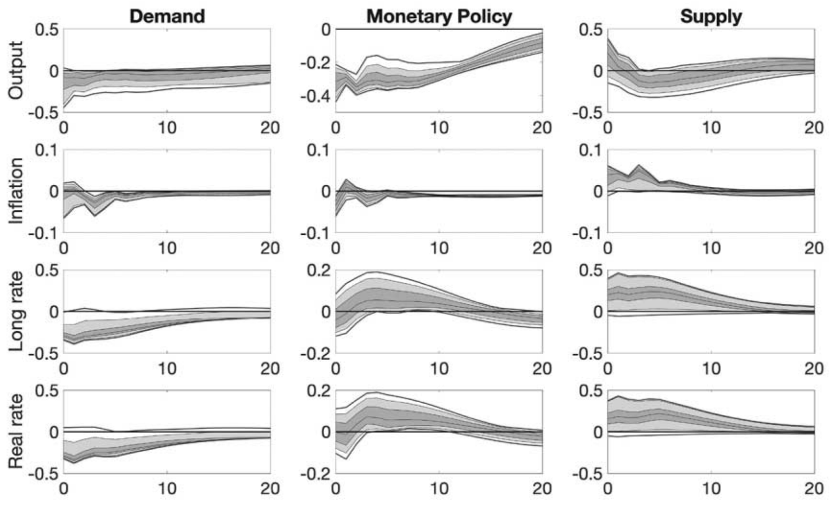

Fig. 6. Impulse responses $\partial Y_{t+h}/\partial\varepsilon_t$. Identified sets and quantiles of truncated prior (= posterior). The outer dark gray lines show the boundaries of the identified sets. The dark (light) error bands show the equal-tail 68% (95%) posterior/prior credible intervals.

the impulse responses show the response of, say $Y_{i,t+h}$, measured in Y_i-units, to a one standard deviation shock in $\varepsilon_{j,t}$. But how large is one standard deviation in $\varepsilon_{j,t}$; for example, how large is a one standard deviation "monetary policy shock" or a one standard deviation "demand shock"? As argued elsewhere (see Stock and Watson 2016), a more interpretable normalization is the "unit-effect" normalization that imposes $\partial Y_{k,t}/\partial\varepsilon_{j,t} = 1$ for some variable $Y_{k,t}$, so that $\varepsilon_{j,t}$ is measured in units of the observed data $Y_{k,t}$. As an example, if $\varepsilon_{j,t}$ is a monetary policy shock, then it is natural to measure its magnitude in terms of basis points of the short-term interest rate. In this case, impulse responses measure the response of macroeconomic variables to a monetary policy shock that, for example, raises short-term interest rates by 100 (or 25) basis points. As a matter of arithmetic, the unit-effect normalized impulse responses can be computed from the unit-standard-deviation normalized impulse responses as the ratio:

$$\frac{\partial Y_{i,t+h}}{\partial\tilde{\varepsilon}_{j,t}} = \frac{\frac{\partial Y_{i,t+h}}{\partial\varepsilon_{j,t}}}{\frac{\partial Y_{k,t}}{\partial\varepsilon_{j,t}}},$$

where $\tilde{\varepsilon}_{j,t}$ is measured in units of $Y_{k,t}$ and $\varepsilon_{j,t}$ is measured in standard deviation units. In point-identified models, this nonlinearity means that care must be taken in computing standard errors for estimated IRFs when translating from one normalization to another. In sign-identified SVARs, the translation is potentially more serious, as for example the identified set for $\partial Y_{k,t}/\partial\varepsilon_{j,t}$ may include values equal to or close to zero. Looking at figure 3, for example, what is the implied effect on output of monetary policy shock that raises interest rates by 25 basis points?

My second comment advertises the work of Wolf (2018), who studies the points in the identified set for some well-known examples of sign-identified SVARs. Wolf is motivated by the observation that only one point in the identified set shows the true impulse response, that is, the response of $Y_{i,t+h}$ to the true structural shock $\varepsilon_{j,t}$. The other points in the identified set implicitly show the response of $Y_{i,t+h}$ to other linear combinations of the structural shocks that also satisfy the sign restrictions. Wolf uses structural models to determine these linear combinations that "masquerade" (his term) for the true structural shock. He uses this to gauge the identifying power of sign restrictions in different classes of models.

Finally, what does all this mean for empirical practice? Sign restrictions can yield valuable identifying information for sorting out cause and effect in macroeconomics. This is particularly the case when these sign restrictions are used in tandem with more traditional equality restrictions (see Arias, Rubio-Ramirez, and Waggoner 2018; Wolf 2018). The narrow identified sets for the Debortoli et al. model (see fig. 3) is an example of the identifying power of these restrictions. That said, the nature of set identification makes inference nonstandard and/or difficult to interpret. One approach is to conduct inference on the identified sets using frequentist methods (e.g., Granziera, Moon, and Schorfheide 2018). This approach usefully summarizes what the data, together with the sign restrictions, have to say about the value of impulse responses. One may be tempted to move beyond these identified sets to produce point estimates, error bands, and so forth, but with set identification this requires Bayes methods, and the associated point estimates, error bands, and so forth merely reproduce the researchers prior over the identified set. Here, "good" priors lead to good inference and conversely for bad priors. Sorting out the good from the bad requires careful presentation and justification for the prior actually used, a point forcefully and convincingly made in theory and practice in Baumeister and Hamilton (2015, 2019). In this regard, the kinds of flat (Haar) priors made on the rotation matrix R in equation (1) seem counterproductive. These flat priors on R produce informative priors on the impulse responses (see figs. 4–6 for examples in a version of the Debortoli et al. SVAR) in ways that are difficult to know a priori, and in any event are rarely, if ever, reported or justified.

Endnote

Author email address: Watson (mwatson@princeton.edu). For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see https://www.nber.org/chapters/c14243.ack.

References

Arias, Jonas, Juan Rubio-Ramirez, and Daniel Waggoner. 2018. "Inference Based on SVARs Identified with Sign and Zero Restrictions." *Econometrica* 86 (2): 685–720.

Baumeister, Christine, and James D. Hamilton. 2015. "Sign Restrictions, Vector Autoregressions, and Useful Prior Information." *Econometrica* 83 (5): 1963–99.

———. 2019. "Structural Interpretation of Vector Autoregressions with Incomplete Information: Revisiting the Role of Oil Supply and Demand Shocks." *American Economic Review* 109 (5): 1873–910.

Granziera, Eleonora, Hyungsik Roger Moon, and Frank Schorfheide. 2018. "Inference for VARs Identified with Sign Restrictions." *Quantitative Economics* 9 (3): 1087–121.

Manski, Charles F. 2003. *Partial Identification of Probability Distributions*. New York: Springer.

Stock, James H., and Mark W. Watson. 2016. "Factor Models and Structural Vector Autoregressions in Macroeconomics." In *Handbook of Macroeconomics*, Vol. 2A, ed. John B. Taylor and Harald Uhlig. Amsterdam: North Holland.

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

Wolf, Christian K. 2018. "SVAR (Mis-)Identification and the Real Effects of Monetary Policy." Manuscript, Princeton University.