Inference in Structural Vector Autoregressions When the Identifying Assumptions are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations*

Christiane Baumeister † James D. Hamilton ‡

October 9, 2015
Revised February 17, 2018

Abstract

Reporting credible sets or error bands for structural vector autoregressions that are only set identified is a very common practice. However, unless the researcher is persuaded on the basis of prior information that some parameter values are more plausible than others, this common practice has no formal justification. When the role and reliability of prior information is defended, Bayesian posterior probabilities can be used to form an inference that incorporates doubts about the identifying assumptions. We illustrate how prior information can be used about both structural coefficients and the impacts of shocks, and propose a new distribution, which we call the asymmetric t distribution, for incorporating prior beliefs about the signs of equilibrium impacts in a nondogmatic way. We apply these methods to a three-variable macroeconomic model and conclude that monetary policy shocks were not the major driver of output, inflation, or interest rates during the Great Moderation.

Keywords: structural vector autoregressions, set identification, monetary policy, impulse-response functions, historical decompositions, model uncertainty, informative priors

---

*We thank Marek Jarocinski, Lam Nguyen and Tao Zha for helpful comments. An earlier version of this paper was circulated under the title "Optimal Inference about Impulse-Response Functions and Historical Decompositions in Incompletely Identified Structural Vector Autoregressions."

†Department of Economics, University of Notre Dame, cjsbaumeister@gmail.com

‡Department of Economics, University of California, San Diego, jhamilton@ucsd.edu

1 Introduction.

A common approach to analyzing dynamic economic relations relies on linear structural models of the form

$$\mathbf{A}\mathbf{y}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u}_t \tag{1}$$

for $\mathbf{y}_t$ an $(n \times 1)$ vector of observed variables at date t, $\mathbf{A}$ an $(n \times n)$ matrix summarizing their contemporaneous structural relations, $\mathbf{x}_{t-1}$ a $(k \times 1)$ vector (with $k = mn + 1$) containing a constant and m lags of $\mathbf{y}$ ($\mathbf{x}'_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, ..., \mathbf{y}'_{t-m}, 1)'$), and $\mathbf{u}_t$ white noise with variance matrix $\mathbf{D}$. Let $\boldsymbol{\theta}$ denote the vector consisting of the unknown elements in $\mathbf{A}, \mathbf{B}$, and $\mathbf{D}$. If we knew the value of $\boldsymbol{\theta}$, the structural model would allow us to make statements about the dynamic effects of the structural shocks $\mathbf{u}_t$.

The reduced form of this structural model is a vector autoregression (VAR):

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2}$$

$$\mathbf{\Phi} = \mathbf{A}^{-1}\mathbf{B} \tag{3}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{A}^{-1}\mathbf{u}_t. \tag{4}$$

The parameters of the VAR, $\mathbf{\Phi}$ and $\mathbf{\Omega} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t)$, can readily be estimated by OLS regressions. The problem is that in the absence of additional information about the structural model, there is no unique mapping from the VAR parameters $\mathbf{\Phi}$ and $\mathbf{\Omega}$ to the structural parameter $\boldsymbol{\theta}$ of interest.[1]

[1] For example, if we considered the structural shocks to be mutually uncorrelated with variance normalized to unity ($\mathbf{D} = \mathbf{I}_n$), there would be n^2 unknown elements of $\mathbf{A}$ but only $n(n+1)/2 < n^2$ elements of $\mathbf{\Omega}$.

Identification requires drawing on additional information about the structural model. For example, $\boldsymbol{\theta}$ would be identified if we knew that $\mathbf{A}$ is lower triangular and $\mathbf{D}$ is diagonal, corresponding to the popular Cholesky or recursive identification scheme. However, such restrictions are rarely completely convincing. For this reason, it has recently become quite common to perform structural analysis relying on less than a complete set of identifying assumptions, for example, knowing only the signs of the effects of certain shocks, an approach pioneered by Canova and De Nicoló (2002) and Uhlig (2005). The most popular algorithm for doing this was developed by Rubio-Ramírez, Waggoner and Zha (2010). Their approach: (1) generates a draw for $\boldsymbol{\Omega}$ from the posterior distribution of the reduced-form covariance matrix resulting from an uninformative Normal-inverse-Wishart prior for the reduced-form parameters, (2) finds the Cholesky factorization $\boldsymbol{\Omega} = \mathbf{PP}'$, (3) draws an orthonormal matrix $\mathbf{Q}$ from a Haar-uniform distribution, (4) proposes $\mathbf{PQ}$ as a candidate draw for the value for the impact matrix, and (5) keeps the draw if it satisfies the sign restrictions.

Researchers typically report the median of the set of accepted values as the most plausible estimate of structural objects of interest and bands around the median containing 68% or 90% of the accepted values as if they were credible sets or error bands. Online Appendix C provides a list of close to a hundred representative studies that have all done this.

To illustrate our concern with this method, consider a simple 3-variable macroeconomic model based on the output gap, inflation, and fed funds rate.[2] Suppose our interest is in what happens to the output gap s quarters after a monetary policy contraction that raises

[2] Details of the data are provided in Section 3 below.

the fed funds rate by 25 basis points. We calculated the answer to this question using the Rubio-Ramírez, Waggoner and Zha (2010) algorithm with one important departure from usual practice– we did not impose any sign restrictions at all, but simply kept every single draw for $\mathbf{PQ}$ from step (4) as a potential answer to the question.[3] Panel A of Figure 1 shows the median along with a band that contains 68% of the generated draws at each horizon. The graph raises a troubling question. The error bands seem to suggest that the impact of a monetary shock is likely to be somewhere between a 0.5% decrease and a 1.0% increase in output. How can we claim to have any confidence in such a statement if we have not made any assumptions?

Panel B clarifies what is going on by plotting the histogram of the draws for the effect at horizon $s = 0$. Baumeister and Hamilton (2015, equation (34)) characterized this distribution analytically. For any given $\mathbf{\Omega}$ this turns out to be a Cauchy distribution with centrality parameter determined by the correlation between the reduced-form residuals for output and the interest rate. When that correlation is positive (as it is for this data set), the centrality parameter is positive, giving rise to the impression that a positive effect is somehow more plausible than a negative effect.[4]

The randomness of the distribution in Panel B comes from two sources. The first is a distribution across different draws of $\mathbf{\Omega}$ from the Normal-inverse-Wishart posterior. The

[3] See Appendix A for details of the algorithm.

[4] If the correlation in the data were zero, the implied error bands would be symmetric around zero. Since the Normal-inverse-Wishart prior anticipates positive and negative correlations as equally likely before seeing the data, the prior itself implies symmetric error bands. Our key concern is not whether the bands are asymmetric, but rather whether there is any basis in the data for thinking that points within the reported error bands are in any sense more plausible than those outside the error bands.

randomness of this first distribution comes from uncertainty about the value of $\mathbf{\Omega}$ given that we have only observed a finite number of observations on $\mathbf{y}_t$. If we had an infinite number of observations on $\mathbf{y}_t$, this distribution would collapse to a point mass at the maximum likelihood estimate $\hat{\mathbf{\Omega}}$. The second source of randomness is the distribution across different draws of the orthonormal matrix $\mathbf{Q}$. The randomness of this second distribution is something introduced by the algorithm itself and has nothing to do with the data.

Panels C and D of Figure 1 clarify the respective contributions of these two sources of randomness by shutting down the first one altogether. To generate these panels, we simply fixed $\mathbf{\Omega}$ at the maximum likelihood estimate $\hat{\mathbf{\Omega}}$ with Cholesky factorization $\hat{\mathbf{\Omega}} = \hat{\mathbf{P}}\hat{\mathbf{P}}'$, and generated 50,000 draws of $\mathbf{Q}$, keeping every single draw of $\hat{\mathbf{P}}\mathbf{Q}$. The randomness in panels C and D comes only from the distribution of $\mathbf{Q}$ and has nothing to do with uncertainty from the data. Panel D is virtually identical to Panel B. This makes clear that the randomness in Panel B comes almost entirely from the randomness introduced by $\mathbf{Q}$.

Nevertheless, *every* draw for $\hat{\mathbf{P}}\mathbf{Q}$ *by construction* fits the observed data $\hat{\mathbf{\Omega}}$ equally well. If we let h denote the magnitude on the horizontal axis in panel D, for *any* $h \in (-\infty, +\infty)$, there exists a value of $\mathbf{Q}$ for which that value of h would be perfectly consistent with the observed $\hat{\mathbf{\Omega}}$. If we claim (as the median line and error bands in Panel C seem to) that some values of h are more plausible than others, what exactly is the basis for that conclusion? Since there is no basis in the data for choosing one value $\hat{\mathbf{P}}\mathbf{Q}$ over any other, any plot highlighting the median or 68% credibility sets of the generated $\hat{\mathbf{P}}\mathbf{Q}$ is relying on an implicit Bayesian prior distribution, according to which some values of h were regarded *a priori* to be more plausible than others. If that is the researchers' intention, then their use of 68% credibility bands would be fine. But none of the papers listed in Appendix C openly acknowledge that a key reason that certain outcomes appear to be ruled out by their credibility bands is because the researcher simply ruled them out a priori even though they are perfectly consistent with all the observed data.[5]

In the last two panels of Figure 1 we perform a slightly more conventional application of the method and only keep the draw from step (4) of the Rubio-Ramírez, Waggoner and Zha (2010) algorithm if it implies that a contractionary monetary shock raises the fed funds rate and lowers output. Panel E looks more like something one might try to publish. But Panel F clarifies that it is simply a truncation of the distribution in panels B or D, numerically shifting the median and all quantiles of the distribution down. There is again no basis in the data for choosing one point in this distribution, or some subset of this distribution, over any other.[6]

There are two ways one can try to do this correctly. One is to remain faithful to the idea that we know absolutely nothing besides the sign restrictions. If this is the goal, then a researcher should not be reporting point estimates or quantiles of a distribution, but should instead describe the complete set of values in which a parameter of interest $\boldsymbol{\theta}$ could be, known as the "identified set." For Panel B the identified set is the real line,

[5] Arias, Rubio-Ramírez, and Waggoner (forthcoming) acknowledged that their prior over $(\boldsymbol{\Phi}, \boldsymbol{\Omega}, \mathbf{Q})$ implies a prior over $(\mathbf{A}, \mathbf{D}, \mathbf{B})$, but did not mention the fact that the bounds for 68% error bands for functions of $(\mathbf{A}, \mathbf{D}, \mathbf{B})$ that emerge from their procedure depend fundamentally on giving an informative role to the distribution used to generate $\mathbf{Q}$.

[6] Song (2014) noted that if we have a minimax loss function, it might be reasonable to report the midpoint of the identified set. However, this is not the same as the median draw and moreover does not exist in the two examples in Figure 1 in which the identified set is unbounded.

while for Panel F it is the set of all negative real numbers.[7] But the plotted error bands in Panels A and E give the misleading impression that we somehow know more than this on the basis of the sign restrictions alone. As one imposes additional restrictions the identified set may become bounded and therefore potentially interesting to report, though often difficult to characterize analytically.[8] Moon et al. (forthcoming) and Gafarov, Meier and Montiel Olea (forthcoming) developed algorithms to estimate the identified set or its bounds using a frequentist approach, while Kline and Tamer (2016) discussed Bayesian posterior inference about the identified set in a general context. Giacomini and Kitagawa (2015) and Gafarov, Meier and Montiel Olea (2016) noted that the identified set could be interpreted and calculated as robust Bayesian posterior inference across the set of all possible Bayesian prior distributions.

But the fact that nearly a hundred prominent studies listed in Appendix C have summarized results based on a strict subset of the identified set suggests to us a need to clarify the conditions under which such a practice could be justified. In this paper we demonstrate that such a justification could come from Bayesian optimal statistical decision theory. Suppose we were willing to let our inference be guided not just by prior information about signs but also about magnitudes. For example, it seems pretty unlikely that a 25-basis-point interest-rate hike would lead to a decline in output as large as 1% and even less likely that

[7] For some questions that the researcher might ask the identified set may be bounded by definition. For example, the Cauchy-Schwartz Inequality implies that the absolute value of the effect of a one-standard deviation shock to structural equation j on variable i cannot exceed the unconditional standard deviation of the innovation to the reduced-form residual for variable i; see Baumeister and Hamilton (2015, p. 1973).

[8] See Amir-Ahmadi and Uhlig (2015) for an analysis of how the size of the identified set can shrink with additional restrictions.

it could lead to a 5% decline. There is a sensible statistical inference in such a setting that comes from weighting the different elements in the identified set by their prior plausibility. This implicitly is what researchers are doing with existing methods, with one very important difference– they do not claim that the distribution in Panel B is a reasonable representation of prior information or even acknowledge that prior information like this has had an influence on the summary statistics they report.

As noted by Baumeister and Hamilton (2015, 2017), using Bayesian priors to assign plausibility to different magnitudes within the identified set can also be regarded as a strict generalization of full identification. For example, Cholesky identification can be viewed as a dogmatic prior in which certain elements of $\mathbf{A}$ are known with certainty to be zero. This can be generalized with an informative prior that those elements of $\mathbf{A}$ are likely to be close to zero, though we're not completely certain they are exactly zero. The model is then no longer formally identified, but the researcher can nevertheless report a valid posterior distribution in which uncertainty about the identifying assumptions themselves (in the form of a probability of how likely it is that elements of $\mathbf{A}$ could be a certain distance from zero) is formally and correctly incorporated in statements of what is plausible having seen the data.

The idea of using informative Bayesian priors as a softer form of identification is not new. It dates back to Drèze (1974) and Drèze and Morales (1976), who made the point in the context of traditional simultaneous equations systems, of which structural vector autoregressions (SVARs) might be viewed as a special case. However, to our knowledge there have been no practical applications of this idea to SVARs prior to ours, suggesting some value in spelling out exactly how it can be done in practice.

In Section 2 we demonstrate that for typical loss functions, the optimal estimate of a structural impulse-response function in an unidentified model with an informative prior can be obtained from the Bayesian posterior mean or posterior median, calculated pointwise for each horizon. This provides a formal justification for the procedure typically adopted by users of sign-restricted SVARs, *provided they are willing to acknowledge the role played by an informative prior.* Our analysis also addresses the concern raised by Fry and Pagan (2011) that the posterior median impulse-response function from a sign-restricted SVAR is not consistent with any fixed value for $\boldsymbol{\theta}$. We document formally that from the point of view of statistical decision theory, the optimal inference about the impulse response at two different horizons *should not* be based on the same value of $\boldsymbol{\theta}$, justifying the straightforward approach that most researchers want to use. Our results further contribute to the discussions by Sims and Zha (1999), Lütkepohl (1990, 2005), Jordà (2009) and Montiel Olea and Plagborg-Møller (2017) on how to estimate and report uncertainty about impulse-response functions.

We further show that analogous results hold for calculating the contributions of individual structural shocks to a given historical episode of interest. To our knowledge, every application of sign-restricted SVARs prior to ours simply plotted the median paths for historical decompositions with no error bands, despite the fact that under their acknowledged assumptions the only valid inference is about intervals rather than a point like the median within the interval. The explanation appears to be that researchers were not sure how to calculate error bands or even how they should be interpreted. In this paper we show that it is again straightforward to characterize both an optimal point estimate and posterior confidence in this estimate as long as the prior used in the analysis is explicit.

Section 3 illustrates these methods using a three-variable macroeconomic model. It is common to conduct macroeconomic analysis with models in which parameters are not estimated at all, but rather are calibrated on the basis of plausible values. We show how information like this can be used to motivate a prior distribution for $\boldsymbol{\theta}$ that would allow a researcher to interpret the contribution of monetary policy to the observed behavior of output, inflation and interest rates even though the analyst has doubts about the identifying assumptions. We further show how information about either the structural coefficients in $\mathbf{A}$ or the equilibrium impacts of structural shocks ($\mathbf{A}^{-1}$) can be used to help reach structural conclusions. We find that given uncertainty about the model itself, the data are not informative about the slope of the Phillips Curve but contain some useful information about the effect of inflation on aggregate demand and Taylor Rule parameters governing the response of the Federal Reserve to the output gap and inflation. Overall, after seeing the data, a researcher would be more confident that a monetary contraction lowers output and inflation. However, we find no strong evidence of an effect on output lasting beyond a few quarters, and monetary policy shocks typically make only a modest contribution to economic fluctuations.

Section 4 demonstrates that our key conclusions do not change if we were to throw out completely any one of the individual sources of information from which our prior is built. Section 5 briefly concludes.

2 Inference in the presence of doubts about the identifying assumptions.

Let $\mathbf{Y}_T = (\mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_T')'$ denote the vector of observed data. Given a distributional assumption for the structural shocks in equation (1), the likelihood function $p(\mathbf{Y}_T|\boldsymbol{\theta})$ can be calculated. For example, if $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{D})$,

$$\begin{aligned} p(\mathbf{Y}_T|\boldsymbol{\theta}) \;&=\; (2\pi)^{-Tn/2}|\det(\mathbf{A}(\boldsymbol{\theta}))|^T|\mathbf{D}(\boldsymbol{\theta})|^{-T/2} \times \\ &\exp\left[-(1/2)\textstyle\sum_{t=1}^{T}(\mathbf{A}(\boldsymbol{\theta})\mathbf{y}_t - \mathbf{B}(\boldsymbol{\theta})\mathbf{x}_{t-1})'\mathbf{D}(\boldsymbol{\theta})^{-1}(\mathbf{A}(\boldsymbol{\theta})\mathbf{y}_t - \mathbf{B}(\boldsymbol{\theta})\mathbf{x}_{t-1})\right] \end{aligned} \quad (5)$$

where $|\det(\mathbf{A})|$ denotes the absolute value of the determinant of $\mathbf{A}$. Given a prior distribution $p(\boldsymbol{\theta})$, the Bayesian posterior distribution is

$$p(\boldsymbol{\theta}|\mathbf{Y}_T) = \frac{p(\mathbf{Y}_T|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{Y}_T|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (6)$$

A suggested class of priors $p(\boldsymbol{\theta})$ and algorithm for generating draws $\{\mathbf{A}^{(\ell)}, \mathbf{D}^{(\ell)}, \mathbf{B}^{(\ell)}\}_{\ell=1}^N$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y}_T)$ that can handle most applications of interest is described in Section 3.

From the reduced-form VAR in (2) we can calculate the nonorthogonalized impulse-response function at horizon s,

$$\boldsymbol{\Psi}_s = \frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'}, \quad (7)$$

by iteration on equation (2) (see for example Hamilton, 1994, p. 260). In particular,

$\boldsymbol{\Psi}_0 = \mathbf{I}_n$ and $\boldsymbol{\Psi}_1$ is given by the first n rows and n columns of $\mathbf{A}^{-1}\mathbf{B}$.[9]

2.1 Inference and credibility sets for impulse-response functions.

Typically researchers are interested in structural objects such as the response of the vector $\mathbf{y}_{t+s}$ to a one-off increase in the jth structural disturbance u_{jt} at time t. For $s = 0$ the answer to this question is given by the jth column of $\mathbf{A}^{-1}$ and for higher s can be found from the jth column of

$$\mathbf{H}_s = \boldsymbol{\Psi}_s \mathbf{A}^{-1}. \tag{8}$$

Let $h_{ij}^s(\boldsymbol{\theta})$ be the value for the effect of the jth structural shock at time t (u_{jt}) on the ith observed variable $y_{i,t+s}$ at time $t+s$.

Suppose our interest is not just in the value h_{ij}^s for some particular value s, but we care instead about the entire function as represented by the $(S \times 1)$ vector $\mathbf{h}_{ij}(\boldsymbol{\theta}) = (h_{ij}^0(\boldsymbol{\theta}), h_{ij}^1(\boldsymbol{\theta}), ..., h_{ij}^{S-1}(\boldsymbol{\theta}))'$. According to Bayesian statistical decision theory, the estimate we report for the $(S \times 1)$ vector should be the value $\hat{\mathbf{h}}_{ij}$ that minimizes the expected loss associated with our choice of $\hat{\mathbf{h}}_{ij}$ where this expectation is taken with respect to the posterior distribution of $\boldsymbol{\theta}$:

$$\hat{\mathbf{h}}_{ij} = \underset{\min \tilde{\mathbf{h}}_{ij}}{\arg} \int g(\mathbf{h}_{ij}(\boldsymbol{\theta}), \tilde{\mathbf{h}}_{ij}) p(\boldsymbol{\theta}|\mathbf{Y}_T) d\boldsymbol{\theta}. \tag{9}$$

Here $g(\mathbf{h}_{ij}, \hat{\mathbf{h}}_{ij})$ is a loss function summarizing how upset we would be if our estimate of the

[9] $\boldsymbol{\Psi}_s$ can equivalently be calculated from the top-left $(n \times n)$ block of

$$\begin{bmatrix} \multicolumn{2}{c}{\underset{n\times(nm)}{\boldsymbol{\Phi}_1}} \\ \underset{(m-1)n\times(m-1)n}{\mathbf{I}_{(m-1)n}} & \underset{(m-1)n\times n}{\mathbf{0}} \end{bmatrix}^s$$

with $\boldsymbol{\Phi}_1$ the first n rows and $k-1$ columns of $\mathbf{A}^{-1}\mathbf{B}$.

function is $\hat{\mathbf{h}}_{ij}$ but the true value is $\mathbf{h}_{ij}$. A leading example is the quadratic loss function:

$$g(\mathbf{h}_{ij}(\boldsymbol{\theta}), \hat{\mathbf{h}}_{ij}) = [\hat{\mathbf{h}}_{ij} - \mathbf{h}_{ij}(\boldsymbol{\theta})]'\mathbf{W}[\hat{\mathbf{h}}_{ij} - \mathbf{h}_{ij}(\boldsymbol{\theta})]. \tag{10}$$

Here $\mathbf{W}$ is a positive definite $(S \times S)$ weighting matrix summarizing our loss function. For example, those elements of $\mathbf{h}_{ij}$ about which we care most would be associated with larger values along the diagonal of $\mathbf{W}$, while the (r, s) off-diagonal term summarizes how an error in predicting term r changes the marginal benefit of getting term s correct. The loss function allows for interaction terms to capture how much we care about getting different elements of the impulse-response function correct.

Let $\mathbf{h}_{ij}^*$ denote the posterior mean of $\mathbf{h}_{ij}$:

$$\mathbf{h}_{ij}^* = \int \mathbf{h}_{ij}(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Y}_T)d\boldsymbol{\theta}.$$

Note that this expression explicitly takes into account the fact that the S elements of $\mathbf{h}_{ij}$ are all functions of the same vector $\boldsymbol{\theta}$, and the Bayesian posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y}_T)$ incorporates the common economic structure and common basis for statistical inference for all the different s. Nevertheless, this expression is calculated simply by finding the posterior mean for each individual h_{ij}^s in isolation and collecting these in a vector. It turns out[10]

[10] Notice that

$$\begin{aligned}
& \int [\hat{\mathbf{h}} - \mathbf{h}(\boldsymbol{\theta})]'\mathbf{W}[\hat{\mathbf{h}} - \mathbf{h}(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta} \\
= & \int [\hat{\mathbf{h}} - \mathbf{h}^* + \mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]\mathbf{W}[\hat{\mathbf{h}} - \mathbf{h}^* + \mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta} \\
= & [\hat{\mathbf{h}} - \mathbf{h}^*]'\mathbf{W}[\hat{\mathbf{h}} - \mathbf{h}^*] + 2[\hat{\mathbf{h}} - \mathbf{h}^*]'\mathbf{W}\int [\mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta} \\
& + \int [\mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]'\mathbf{W}[\mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta} \\
= & [\hat{\mathbf{h}} - \mathbf{h}^*]'\mathbf{W}[\hat{\mathbf{h}} - \mathbf{h}^*] + \int [\mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]'\mathbf{W}[\mathbf{h}^* - \mathbf{h}(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}
\end{aligned}$$

that the vector of posterior means is also the solution to (9): $\hat{\mathbf{h}}_{ij} = \mathbf{h}^*_{ij}$. In other words, the point-by-point posterior means of each individual element of the impulse-response function represent the values we should use even when our interest is in the entire function $\mathbf{h}_{ij}$ regardless of the value of the weights $\mathbf{W}$. Note that this optimal estimate can easily be calculated pointwise from the set of posterior draws, namely

$$\hat{\mathbf{h}}_{ij} = \begin{bmatrix} N^{-1}\sum_{\ell=1}^{N} h^0_{ij}(\boldsymbol{\theta}^{(\ell)}) \\ N^{-1}\sum_{\ell=1}^{N} h^1_{ij}(\boldsymbol{\theta}^{(\ell)}) \\ \vdots \\ N^{-1}\sum_{\ell=1}^{N} h^{S-1}_{ij}(\boldsymbol{\theta}^{(\ell)}) \end{bmatrix}.$$

Ninety-five percent posterior credibility regions can be calculated from the upper and lower 2.5% quantiles of $h^s_{ij}(\boldsymbol{\theta}^{(\ell)})$.[11]

Alternatively, if our loss function is instead

$$g(\mathbf{h}_{ij}, \hat{\mathbf{h}}_{ij}) = \omega_0 \left| h^0_{ij} - \hat{h}^0_{ij} \right| + \omega_1 \left| h^1_{ij} - \hat{h}^1_{ij} \right| + \cdots + \omega_{S-1} \left| h^{S-1}_{ij} - \hat{h}^{S-1}_{ij} \right|$$

for any set of positive weights $\{\omega_s\}_{s=0}^{S-1}$, it is not hard to show[12] that element s of the optimal estimate $\hat{\mathbf{h}}_{ij}$ is the posterior median of $h^s_{ij}(\boldsymbol{\theta}^{(\ell)})$.[13]

which is minimized with respect to $\hat{\mathbf{h}}$ by setting $\hat{\mathbf{h}} = \mathbf{h}^*$.

[11] Note that we do not propose use of such intervals for purposes of making a statistical decision, but instead simply as a convenient visual device for summarizing an important feature of the posterior distribution.

[12] For this case we have

$$\frac{\partial}{\partial \hat{h}^s_{ij}} \int g(\mathbf{h}_{ij}, \hat{\mathbf{h}}_{ij}) p(\boldsymbol{\theta}|\mathbf{Y}_T) = \omega_s \left\{ -\Pr\left[h^s_{ij}(\boldsymbol{\theta}) > \hat{h}^s_{ij} | \mathbf{Y}_T \right] + \Pr\left[h^s_{ij}(\boldsymbol{\theta}) \leq \hat{h}^s_{ij} | \mathbf{Y}_T \right] \right\}$$

which equals zero when $\hat{h}^s_{ij}$ satisfies $\Pr\left[h^s_{ij}(\boldsymbol{\theta}) \leq \hat{h}^s_{ij} | \mathbf{Y}_T \right] = 0.5$.

[13] That is, for each individual i, j, and s, we order the draws such that $h^s_{ij}(\boldsymbol{\theta}^{(\ell^*_{i,j,s}+1)}) > h^s_{ij}(\boldsymbol{\theta}^{(\ell^*_{i,j,s})})$ and take $\hat{h}^s_{ij} = h^s_{ij}(\boldsymbol{\theta}^{(\ell^*_{i,j,s})})$ for $\ell^*_{i,j,s} = N/2$.

To relate this conclusion to the Fry and Pagan (2011) critique, consider the special case of a univariate AR(1), $y_t = \theta y_{t-1} + \varepsilon_t$. Suppose that our object of interest is the impulse response at horizons 1 and 2:

$$\mathbf{h}(\theta) = \begin{bmatrix} \partial y_{t+1}/\partial \varepsilon_t \\ \partial y_{t+2}/\partial \varepsilon_t \end{bmatrix} = \begin{bmatrix} \theta \\ \theta^2 \end{bmatrix}.$$

Suppose for illustration that the posterior distribution is Gaussian: $\theta|\mathbf{Y}_T \sim N(\mu, \sigma^2)$. Then

$$\mathbf{h}^* = \begin{bmatrix} E(\theta|\mathbf{Y}_T) \\ E(\theta^2|\mathbf{Y}_T) \end{bmatrix} = \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix}. \tag{11}$$

It might seem odd at first that the optimal estimate of the second element, $\mu^2 + \sigma^2$, is not the square of the estimate of the first element, μ, given that the second element of $\mathbf{h}$ for any fixed value of θ is always the square of the first. But this difference between the optimal estimates of $\partial y_{t+1}/\partial \varepsilon_t$ and that for $\partial y_{t+2}/\partial \varepsilon_t$ is a necessary implication of Jensen's inequality given that the elements of the impulse-response function are nonlinear functions of the underlying parameter θ. Reporting the estimate of the impulse-response function to be the magnitudes in (11) is the unique optimal solution to (9) given (10), and any estimate of $\mathbf{h}$ other than (11), such as the estimate $\tilde{\mathbf{h}} = (\mu, \mu^2)'$, would result in a higher value for the expected loss than does the vector $\mathbf{h}^*$ given in (11). This is because $\tilde{\mathbf{h}}$ gives a worse estimate of the second element of $\mathbf{h}$ and no better estimate of the first element compared to $\mathbf{h}^*$.

Alternatively, the econometrician might wish to report an estimate of the parameter vector $\boldsymbol{\theta}$ itself. Again to talk about optimality of such an estimate we would need a loss function. For example, with a quadratic loss function,

$$g(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{W} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

the optimal estimate is again the element-by-element posterior mean which we obtain from $N^{-1} \sum_{\ell=1}^{N} \boldsymbol{\theta}^{(\ell)}$.

Some researchers have proceeded as if their loss function for choosing $\hat{\boldsymbol{\theta}}$ is

$$g(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = [\mathbf{h}(\boldsymbol{\theta}) - \mathbf{h}(\hat{\boldsymbol{\theta}})]' \mathbf{W} [\mathbf{h}(\boldsymbol{\theta}) - \mathbf{h}(\hat{\boldsymbol{\theta}})] \tag{12}$$

for $\mathbf{h}(\boldsymbol{\theta})$ the $(n^2 S \times 1)$ vector obtained by stacking the impulse-response vectors $\mathbf{h}_{ij}(\boldsymbol{\theta})$ implied by a given value of $\boldsymbol{\theta}$ on top of each other for $i, j = 1, ..., n$. Unlike (10), the solution $\hat{\boldsymbol{\theta}}$ to this problem will depend on the weights $\mathbf{W}$ and will have the property for the AR(1) example that

$$\mathbf{h}(\hat{\theta}) = \begin{bmatrix} \hat{\theta} \\ \hat{\theta}^2 \end{bmatrix}. \tag{13}$$

Christiano, Eichenbaum, and Evans (2005) proposed constructing estimates of $\boldsymbol{\theta}$ directly from this loss function, and Fry and Pagan (2011) and Inoue and Kilian (2013) argued for the importance of the apparent internal consistency provided by (13). From the perspective of statistical decision theory, which approach is better depends on whether the loss function is taken to be (10) or (12). In most applied studies, the emphasis is usually on estimates of the impulse-response functions $\mathbf{h}$. Indeed, estimates of the parameters $\boldsymbol{\theta}$ are typically never even reported, suggesting that the appropriate loss function is (10) rather than (12). This means that in most cases researchers would likely want to report the pointwise posterior means or pointwise posterior medians of $\mathbf{h}$ rather than some other estimates.

2.2 Inference and credibility sets for historical decompositions.

Another feature in which applied researchers are often interested is the contribution of different structural shocks to particular historical episodes of interest. If we knew the value of $\boldsymbol{\theta}$ we could write the value of $\mathbf{y}_{t+s}$ as a known function of initial conditions at time t plus the reduced-form innovations between $t+1$ and $t+s$ (e.g., Hamilton, 1994, equation [10.1.14])

$$\mathbf{y}_{t+s} = \boldsymbol{\Psi}_0(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t+s-2} + \cdots + \boldsymbol{\Psi}_{s-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t+1} + \mathbf{G}_s(\boldsymbol{\theta})\mathbf{x}_t \tag{14}$$

for $\boldsymbol{\Psi}_s(\boldsymbol{\theta})$ the nonorthogonalized impulse-response matrix in (7) and $\mathbf{G}_s(\boldsymbol{\theta})$ the first n rows of the matrix in footnote 9. Conditional on the observed data $\mathbf{Y}_T$ and on knowing $\boldsymbol{\theta}$ we would also know the value of each structural shock at each date in the sample with certainty:

$$\mathbf{u}_t(\boldsymbol{\theta}, \mathbf{Y}_T) = \mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1}.$$

Using (4) we could then write the contribution of structural shocks between $t+1$ and $t+s$ to the value of $\mathbf{y}_{t+s}$ as

$$\mathbf{H}_0(\boldsymbol{\theta})\mathbf{u}_{t+s}(\boldsymbol{\theta}, \mathbf{Y}_T) + \mathbf{H}_1(\boldsymbol{\theta})\mathbf{u}_{t+s-1}(\boldsymbol{\theta}, \mathbf{Y}_T) + \cdots + \mathbf{H}_{s-1}(\boldsymbol{\theta})\mathbf{u}_{t+1}(\boldsymbol{\theta}, \mathbf{Y}_T)$$

for $\mathbf{H}_s(\boldsymbol{\theta})$ the matrix in (8). The contribution to the value of $\mathbf{y}_t$ of structural shock j over the most recent s periods is thus given by the $(n \times 1)$ vector

$$\begin{aligned}\boldsymbol{\zeta}_{jts}(\boldsymbol{\theta}, \mathbf{Y}_T) &= \mathbf{H}_0(\boldsymbol{\theta})\left[\mathbf{e}_j \odot \mathbf{u}_t(\boldsymbol{\theta}, \mathbf{Y}_T)\right] + \mathbf{H}_1(\boldsymbol{\theta})\left[\mathbf{e}_j \odot \mathbf{u}_{t-1}(\boldsymbol{\theta}, \mathbf{Y}_T)\right] + \\ &\quad \cdots + \mathbf{H}_{s-1}(\boldsymbol{\theta})\left[\mathbf{e}_j \odot \mathbf{u}_{t-s+1}(\boldsymbol{\theta}, \mathbf{Y}_T)\right]\end{aligned} \tag{15}$$

where $\mathbf{e}_j$ denotes the jth column of $\mathbf{I}_n$ and $\odot$ denotes element-by-element multiplication.

From a Bayesian perspective, the uncertainty about $\boldsymbol{\zeta}_{jts}(\boldsymbol{\theta}, \mathbf{Y}_T)$ conditional on having observed the full sample of data $\mathbf{Y}_T$ is entirely summarized by the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y}_T)$. Thus for example given a quadratic loss function the optimal estimate of the contribution of the jth structural shock to the evolution of $\mathbf{y}$ between dates $t-s+1$ and t is given by

$$\hat{\boldsymbol{\zeta}}_{jts} = N^{-1} \textstyle\sum_{\ell=1}^{N} \boldsymbol{\zeta}_{jts}(\boldsymbol{\theta}^{(\ell)}, \mathbf{Y}_T). \tag{16}$$

A ninety-five percent credibility set for the effect on variable i can be obtained by sorting $\zeta_{ijts}(\boldsymbol{\theta}^{(\ell)}, \mathbf{Y}_T)$ in increasing order for each i, j and reporting the values $\zeta_{ijts}(\boldsymbol{\theta}^{(\ell^*_{ijs})}, \mathbf{Y}_T)$ for $\ell^*_{ijs} = 0.025N$ and $0.975N$.

One advantage of the quadratic over the absolute-value loss function in this case is that both population and sample means have the property that the mean of the sum is the sum of the means. Since the sum over j of the components (15) exactly equals the realized value of $\mathbf{y}_{t+s} - \mathbf{G}_s(\boldsymbol{\theta})\mathbf{x}_t$ for every $\boldsymbol{\theta}$, the sum of the estimated components (16) also exactly matches the observed data.

2.3 Inference and credibility sets for variance decompositions.

It follows from the above analysis of equation (14) that conditional on $\boldsymbol{\theta}$ the s-period-ahead error in forecasting the observable variables can be written as

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} = \mathbf{H}_0(\boldsymbol{\theta})\mathbf{u}_{t+s} + \mathbf{H}_1(\boldsymbol{\theta})\mathbf{u}_{t+s-1} + \mathbf{H}_2(\boldsymbol{\theta})\mathbf{u}_{t+s-2} + \cdots + \mathbf{H}_{s-1}(\boldsymbol{\theta})\mathbf{u}_{t+1}$$

whose mean squared error (MSE) is

$$E[(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'|\boldsymbol{\theta}] = \textstyle\sum_{j=1}^{n} \mathbf{Q}_{js}(\boldsymbol{\theta})$$

$$\mathbf{Q}_{js}(\boldsymbol{\theta}) = d_{jj}(\boldsymbol{\theta}) \sum_{k=0}^{s-1} \mathbf{h}_j(k;\boldsymbol{\theta})\mathbf{h}_j(k;\boldsymbol{\theta})'$$

for $\mathbf{h}_j(k;\boldsymbol{\theta})$ the jth column of $\mathbf{H}_k(\boldsymbol{\theta})$ and $d_{jj}(\boldsymbol{\theta})$ the (j,j) element of $\mathbf{D}$. The contribution of structural shock j to the s-period-ahead MSE of the ith element of $\mathbf{y}_{t+s}$ is given by the (i,i) element of $\mathbf{Q}_{js}(\boldsymbol{\theta})$. An estimate of this magnitude could be obtained from the posterior mean or median across draws of $\boldsymbol{\theta}^{(\ell)}$, $\ell = 1, ..., N$. Again an advantage of the posterior mean is that the estimate of the sum across j of the contributions of individual shocks will equal by construction the estimate of the total s-period-ahead MSE for every s.

3 Bayesian inference in a 3-variable macro model.

Here we illustrate these methods using a commonly studied three-variable macroeconomic model.[14] The three quarterly variables are summarized by the vector $\mathbf{y}_t = (y_t, \pi_t, r_t)'$, where y_t denotes the output gap (100 times the log difference between observed and potential real GDP as estimated by the Congressional Budget Office), π_t the inflation rate (measured by 100 times the year-over-year log change in the personal consumption expenditures deflator), and r_t the nominal interest rate (measured by the average value for the fed funds rate over the quarter).

3.1 Model description.

The system consists of a Phillips Curve,

$$y_t = k^s + \alpha^s \pi_t + [\mathbf{b}^s]' \mathbf{x}_{t-1} + u_t^s, \tag{17}$$

[14] Equations (17)-(19) can be motivated from the 3-variable macro models studied by Rotemberg and Woodford (1997), Lubik and Schorfheide (2004), Del Negro and Schorfheide (2004), Giordani (2004), Benati and Surico (2009), and Rubio-Ramirez, Waggoner, and Zha (2010).

an aggregate demand equation,

$$y_t = k^d + \beta^d \pi_t + \gamma^d r_t + \left[\mathbf{b}^d\right]' \mathbf{x}_{t-1} + u_t^d, \tag{18}$$

and a Taylor Rule for monetary policy,

$$r_t = k^m + \zeta^y y_t + \zeta^\pi \pi_t + [\mathbf{b}^m]' \mathbf{x}_{t-1} + u_t^m, \tag{19}$$

where $\mathbf{x}_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, ..., \mathbf{y}'_{t-m}, 1)'$ and u_t^s denotes a shock to supply, u_t^d the demand shock, and u_t^m the monetary policy shock. We take the number of lags m to be four quarters.[15]

This system will be recognized as a special case of the general framework (1) with

$$\mathbf{A} = \begin{bmatrix} 1 & -\alpha^s & 0 \\ 1 & -\beta^d & -\gamma^d \\ -\zeta^y & -\zeta^\pi & 1 \end{bmatrix}. \tag{20}$$

In the absence of additional information about the elements of $\mathbf{A}$, the model would be unidentified and there would be no basis for drawing conclusions from the data about the effects of monetary policy. The conventional approach is to impose hard restrictions on the elements of $\mathbf{A}$, which can be interpreted as a dogmatic prior. Here we propose instead to use prior beliefs about the underlying economic structure in a less dogmatic fashion, claiming that we do know something about plausible values for these parameters, but do not know any of the values with certainty. We follow Baumeister and Hamilton (2015) in writing the prior $p(\boldsymbol{\theta}) = p(\mathbf{A})p(\mathbf{D}|\mathbf{A})p(\mathbf{B}|\mathbf{A}, \mathbf{D})$ where the functional form of $p(\mathbf{A})$ is completely

[15] Data and code for replicating our results are available at:
http://econweb.ucsd.edu/~jhamilton/BH3_code.zip

unrestricted while those of $p(\mathbf{D}|\mathbf{A})$ and $p(\mathbf{B}|\mathbf{A},\mathbf{D})$ are taken from natural conjugate families to simplify the computational demands. We discuss the priors $p(\mathbf{A})$, $p(\mathbf{D}|\mathbf{A})$ and $p(\mathbf{B}|\mathbf{A},\mathbf{D})$ in the following subsections.

3.2 Prior information about contemporaneous structural coefficients.

It is common in theoretical macroeconomic models to work with a special case of (19) such as

$$r_t - \bar{r} = (1-\rho)\psi^y y_t + (1-\rho)\psi^\pi(\pi_t - \pi^*) + \rho(r_{t-1} - \bar{r}) + u_t^m, \tag{21}$$

where ψ^y and ψ^π describe the Fed's long-run response to output and inflation, π^* is the Fed's long-run inflation target, $\bar{r}$ is the sum of π^* and the long-run real interest rate, and ρ reflects the Fed's desire to implement changes gradually over time. Taylor (1993) proposed values of $\psi^y = 0.5$ and $\psi^\pi = 1.5$. We will represent this structural belief about monetary policy by using a Student t prior for ψ^y with mode at 0.5, scale parameter 0.4, and degrees of freedom $\nu_\psi = 3$, truncated to be positive. The Student t distribution includes as a special case (when $\nu_\psi = 1$) the Cauchy distribution underlying Panel B of Figure 1. A Normal distribution is another special case of the Student t (when $\nu_\psi \to \infty$). Our choice of $\nu_\psi = 3$ represents a bit more confidence than the Cauchy, ensuring that the posterior distribution of ψ^y has a finite variance, though still allowing substantial probabilities in the tails. This density is plotted as a red curve in the lower-left panel of Figure 2. It assigns an 82% prior probability that ψ^y is between 0 and 1 and a 98% prior probability that it is between 0 and 2. For our prior for ψ^π we used a Student t distribution with mode at 1.5, scale parameter

0.4, and 3 degrees of freedom, again truncated to be positive. This density is plotted in red in the bottom middle panel of Figure 2. For the smoothing parameter ρ we follow Lubik and Schorfheide (2004) and Del Negro and Schorfheide (2004) in using a Beta distribution with mean 0.5 and standard deviation 0.2 plotted in the bottom right panel.[16] Priors for these and other contemporaneous parameters are summarized in Table 1.

The joint distribution for the elements in the last row of (20) is thus that of a two-dimensional random variable characterized by

$$\begin{bmatrix} \zeta^y \\ \zeta^\pi \end{bmatrix} = \begin{bmatrix} (1-\rho)\psi^y \\ (1-\rho)\psi^\pi \end{bmatrix} \tag{22}$$

where ρ, ψ^y, and ψ^π have the distributions described in Table 1.

The parameter ρ will also give us prior information about the lagged structural coefficients $\mathbf{b}^m$ in (19). We will describe how we use this information and how the observed dynamics of the variables can help identify ρ separately from ψ^y and ψ^π in Section 3.5. But first we discuss priors for the contemporaneous coefficients in the other structural equations.

The aggregate demand equation (18) is sometimes viewed as the implication of a consumption Euler equation or dynamic IS curve of the form

$$y_t = c^d + \xi y_{t+1|t} - \tau(r_t - \pi_{t+1|t}) + u_t^d \tag{23}$$

where ξ is the weight on the forward-looking component of the IS curve, τ is the intertemporal elasticity of substitution and $y_{t+1|t}$ and $\pi_{t+1|t}$ are one-step-ahead forecasts of output and inflation. One option would be to take a completely specified dynamic stochastic general

[16] Benati (2008) used a mean of 0.5 and standard deviation of 0.25.

equilibrium model, find the rational-expectations solutions $y_{t+1|t} = \boldsymbol{\phi}^{y\prime}\mathbf{x}_t$ and $\pi_{t+1|t} = \boldsymbol{\phi}^{\pi\prime}\mathbf{x}_t$, substitute these expressions into (23), and get values for β^d and γ^d from the contemporaneous coefficients in the resulting equation. These would then characterize the values anticipated for β^d and γ^d as a function of all the parameters of a complete model in a generalization of the technique used to arrive at (22). However, it is much simpler, and more in keeping with the less restrictive and more data-based approach favored in this paper, to draw instead on prior beliefs about the reduced-form coefficients $\boldsymbol{\phi}^y$ and $\boldsymbol{\phi}^\pi$ themselves. Our priors for the reduced-form coefficients are similar to those in Doan, Litterman and Sims (1984) in expecting that a simple AR(1) process probably gives a decent forecast of most economic time series; specifically, $y_{t+1|t} = c^y + \phi^y y_t$ and $\pi_{t+1|t} = c^\pi + \phi^\pi \pi_t$, where our prior expectation is $\phi^y = \phi^\pi = \phi = 0.75$. Substituting these expressions into (23) gives

$$\begin{aligned} y_t &= \mu^d + \phi\xi y_t - \tau(r_t - \phi\pi_t) + u_t^d \\ &= \tilde{\mu}^d - \tilde{\tau}r_t + \tilde{\tau}\phi\pi_t + \tilde{u}_t^d \end{aligned}$$

where $\mu^d = c^d + \xi c^y + \tau c^\pi$ and $\tilde{\tau} = \tau/(1-\phi\xi)$. Benati's (2008) prior for ξ had a mean of 0.5. Benati and Surico's (2009) prior mode was 0.25, whereas Lubik and Schorfheide (2004) imposed $\xi = 1$. A value of $\xi = 2/3$ would imply $\tilde{\tau} = 2\tau$. Many macro models assume an intertemporal elasticity of substitution of $\tau = 0.5$. These considerations led us to use a Student t prior for γ^d in (18) with mode -1, scale parameter 0.4, and 3 degrees of freedom, for which we further impose the sign restriction that γ^d cannot be positive since we are certain that higher interest rates do not stimulate aggregate demand. We likewise use a Student t prior for β^d with mode 0.75. We do not impose a hard sign restriction on β^d

since its sign will depend on the correct specification for forecasts of inflation, about which we do not have strong prior beliefs.

Finally, for the Phillips Curve (17) we follow Lubik and Schorfheide (2004) in using a mode for α^s of 2, implemented again with a Student t distribution now assumed to be positive.

3.3 Prior information about impacts of shocks.

Most applications of sign-restricted SVARs have imposed implicit priors not on the structural coefficients in $\mathbf{A}$ but instead on contemporaneous impacts determined by $\mathbf{H} = \mathbf{A}^{-1}$. Here we show how this can be done using an extension of the algorithm in Baumeister and Hamilton (2015).

There is no reason why prior information about the model could not come from a variety of sources. To illustrate this point with a very simple example, suppose we are interested in a population mean μ of a Gaussian distribution and had earlier observed two independent samples each of size T drawn from this population, the first with sample mean $\bar{y}_1$ and the second with sample mean $\bar{y}_2$. If we were relying on just the first source of information, we would use the prior $p_1(\mu) = \phi(\mu; \bar{y}_1, \sigma^2/T)$ (the Normal density with mean $\bar{y}_1$ and variance σ^2/T evaluated at the point μ). If we were relying on just the second source of information, we would use $p_2(\mu) = \phi(\mu; \bar{y}_2, \sigma^2/T)$. But of course the best procedure is to use both sources of information, and use as our prior for μ the product $p(\mu) = p_1(\mu)p_2(\mu)$. For this example, we can see analytically that this product amounts to a $N\left((\bar{y}_1 + \bar{y}_2)/2, \sigma^2/(2T)\right)$ prior distribution for μ. In more complicated settings, we do not need to solve the problem analytically but can simply take the product of the densities that summarize different independent sources of information. Here we show how information about $\mathbf{H}$ beyond the information previously used about $\mathbf{A}$ can be used to generate a combined prior for $\mathbf{A}$ (expression (30) below) that has the most mass at values of $\mathbf{A}$ that are most consistent with all the various sources of information.

For our model we have

$$\mathbf{A} = \begin{bmatrix} 1 & -\alpha^s & 0 \\ 1 & -\beta^d & -\gamma^d \\ -(1-\rho)\psi^y & -(1-\rho)\psi^\pi & 1 \end{bmatrix} \tag{24}$$

$$\mathbf{H} = \frac{1}{\det(A)}\tilde{\mathbf{H}}$$

$$\det(A) = \alpha^s[1-\gamma^d(1-\rho)\psi^y] - [\beta^d + \gamma^d(1-\rho)\psi^\pi] \tag{25}$$

$$\tilde{\mathbf{H}} = \begin{bmatrix} -[\beta^d + \gamma^d(1-\rho)\psi^\pi] & \alpha^s & \alpha^s\gamma^d \\ \gamma^d(1-\rho)\psi^y - 1 & 1 & \gamma^d \\ -(1-\rho)(\psi^\pi + \beta^d\psi^y) & (1-\rho)(\psi^\pi + \alpha^s\psi^y) & \alpha^s - \beta^d \end{bmatrix}. \tag{26}$$

We have imposed the sign restrictions $\alpha^s > 0$, $\gamma^d < 0$, $\psi^y > 0$, $\psi^\pi > 0$, and $(1-\rho) > 0$. These guarantee the signs of some but not all the elements of $\tilde{\mathbf{H}}$:

$$\mathrm{sign}(\tilde{\mathbf{H}}) = \begin{bmatrix} ? & + & - \\ - & + & - \\ ? & + & ? \end{bmatrix}.$$

In addition, the sign of $\det(\mathbf{A})$ is not determined. The latter is a potential concern because it means that in some allowable regions of the parameter space, elements of $\mathbf{A}^{-1}$ become infinite before flipping signs. If we were to impose the additional restriction that

$$h_1(\boldsymbol{\theta}) = \beta^d + \gamma^d(1-\rho)\psi^\pi < 0, \tag{27}$$

it would guarantee both that $\det(\mathbf{A}) > 0$ and that the (1,1) element of $\tilde{\mathbf{H}}$ is positive, that is, a favorable supply shock raises output and lowers inflation.

In keeping with our theme of relying on partial identifying assumptions that are a strict generalization of previous approaches, we will not impose the inequality (27) dogmatically, but instead will incorporate the prior information that h_1 is probably negative. This probability can be brought arbitrarily close to unity depending on the parameters used to represent the researcher's confidence in the prior information about the signs of impacts.

To do this we introduce a new family of densities that we will refer to as an asymmetric t distribution.[17] Let $\tilde{\phi}_v(x)$ denote the probability density function of a standard Student t variable with ν degrees of freedom evaluated at the point x,[18] and let $\Phi(x)$ denote the cumulative distribution function for a standard $N(0,1)$ variable. Consider a random variable $h \in (-\infty, \infty)$ with the following density, which has location parameter μ_h, scale parameter σ_h, degrees of freedom parameter ν_h and shape parameter λ_h,

$$p(h) = k\sigma_h^{-1}\tilde{\phi}_{v_h}((h-\mu_h)/\sigma_h)\Phi(\lambda_h h/\sigma_h), \tag{28}$$

where k is a constant to make the density integrate to one. The parameter λ_h governs the

[17] The asymmetric t is a straightforward adaptation of the ideas in Azzalini and Capitanio (2003), though to our knowledge the particular density (28) has not appeared previously.

[18] That is,

$$\tilde{\phi}_v(x) = \frac{\Gamma[(\nu+1)/2]}{(\nu\pi)^{1/2}\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

asymmetry of the distribution. If $\lambda_h = 0$, then $\Phi(\lambda_h h/\sigma_h) = 1/2$ for all h and (28) becomes the density of a symmetric Student t variable with location parameter μ_h, scale parameter σ_h, degrees of freedom ν_h, and with the integrating constant $k = 2$. This becomes the further special case of the $N(\mu_h, \sigma_h^2)$ distribution when $\nu_h \to \infty$ and the Cauchy distribution when $\nu_h = 1$. When $\lambda_h > 0$ the density in (28) is positively skewed and when $\lambda_h < 0$ it is negatively skewed. As $\lambda_h \to \infty$, $\Phi(\lambda_h h/\sigma_h)$ goes to 0 for any negative h and goes to 1 for any positive h. Thus when $\lambda_h \to \infty$, (28) becomes a Student t (μ_h, σ_h, ν_h) variable truncated to be positive. When $\lambda_h \to -\infty$, (28) becomes a Student t (μ_h, σ_h, ν_h) truncated to be negative. Thus for example we could include the marginal prior for an impact coefficient constrained to be positive that is implicit in the traditional Haar prior on rotation matrices as a special case when $\lambda_h \to \infty$, $\nu_h = 1$, and μ_h and σ_h are known functions of the reduced-form covariance matrix $\mathbf{\Omega}$.

Our proposed alternative to the implicit Haar prior is instead to rely directly on prior information about structural parameters to specify likely values for a magnitude like h_1. To do this, we drew values for β^d, γ^d, ψ^π, and ρ from the distributions summarized in Table 1 to get a draw for a value for h_1. We used the average value of the simulated h_1 to set the value $\mu_{h_1} = -0.1$ and the standard deviation of the draws to determine $\sigma_{h_1} = 1$. We set $\nu_{h_1} = 3$ and $\lambda_{h_1} = -4$, which strongly nudge the data in the direction of $h_1 < 0$, but still allow a 6.5% chance that $h_1 > 0$. This density is plotted in Panel A of Figure 3.

The proposal is then to take the log of the prior specified in Table 1, namely

$$\log p(\alpha^s) + \log p(\beta^d) + \log p(\gamma^d) + \log p(\psi^y) + \log p(\psi^\pi) + \log p(\rho)$$

and add to it the term

$$\log p(h_1) = \zeta_{h_1}[\log \tilde{\phi}_{v_{h_1}}((h_1 - \mu_{h_1})/\sigma_{h_1}) + \log \Phi(\lambda_{h_1} h_1/\sigma_{h_1})] \tag{29}$$

where $h_1 = \beta^d + \gamma^d(1-\rho)\psi^\pi$ and ζ_{h_1} governs the overall weight put on the prior for h_1. When $\zeta_{h_1} = 0$ the information about h_1 is ignored altogether. We set $\zeta_{h_1} = 1$ for this application.

The algorithm in Baumeister and Hamilton (2015) does not require the prior $p(\mathbf{A})$ to integrate to one since the constant of integration is calculated implicitly through the simulation. Note also that as a result of adding (29), the resulting prior $p(\mathbf{A})$ is no longer independent across the individual elements of $\mathbf{A}$, but includes some joint information about their interaction, favoring combinations of parameters that imply $h_1 < 0$ over those that do not.

Another place we might want to draw on additional information is the (3,3) element of $\tilde{\mathbf{H}}$. Note that the prior as specified so far does not impose that a monetary contraction results in a higher interest rate once equilibrium feedback effects are considered. Here we illustrate how one can use prior information about the plausible magnitude of the effect of monetary policy to assist further with identification. Note from (26) that the response of the output gap to a monetary contraction that raises the fed funds rate by 1 percentage point is

$$h_2 = \frac{\alpha^s \gamma^d}{\alpha^s - \beta^d}.$$

We would expect $h_2 < 0$, but do not impose this, and use instead $\lambda_{h_2} = -2$ as a more modest way of favoring parameter combinations that result in an impact of the expected sign. We set $\mu_{h_2} = -0.3$, a prior expectation that output would fall by 0.3%, with $\sigma_{h_2} = 0.5$, $\nu_{h_2} = 3$, and $\zeta_{h_2} = 1$. This prior is plotted in Panel B of Figure 3. It allows a 6.6% probability that h_2 is in fact positive, that is, that output increases in response to a monetary contraction.[19]

Our baseline specification thus uses

$$
\begin{aligned}
\log p(\mathbf{A}) &= \log p(\alpha^s) + \log p(\beta^d) + \log p(\gamma^d) + \log p(\psi^y) + \log p(\psi^\pi) + \log p(\rho) \\
&\quad + \log p[h_1(\beta^d, \gamma^d, \psi^\pi, \rho)] + \log p[h_2(\alpha^s, \gamma^d, \beta^d)] \qquad (30)
\end{aligned}
$$

for

$$\log p(h_2) = \zeta_{h_2}[\log \tilde{\phi}_{v_{h_2}}((h_2 - \mu_{h_2})/\sigma_{h_2}) + \log \Phi(\lambda_{h_2} h_2/\sigma_{h_2})].$$

Calculations like these of the implied values of $\mathbf{A}^{-1}$ can be a useful check on how parameters can interact in equilibrium, and we recommend this as an additional tool for evaluating the plausibility of prior beliefs. But these calculations also highlight that the equilibrium impacts of shocks can depend in a complicated way on various unknown parameters. It seems preferable to relate beliefs about the likely signs of these impacts to an underlying structural model and acknowledge that we may not know the signs of equilibrium impacts with certainty.

In larger dimensional or more complicated models, it may be tedious to calculate $\det(\mathbf{A})$ and $\tilde{\mathbf{H}}$ analytically as we have done here. But the same basic approach could be implemented entirely numerically. For any given value for $\boldsymbol{\theta}$ one can calculate numerically the determinant

[19] This is also in the spirit of Uhlig (2005), who challenged the conventional wisdom of the real effects of monetary policy.

of $\mathbf{A}$ and the adjoint $\tilde{\mathbf{H}}$ of $\mathbf{A}$. We could thus always calculate $h = \det(\mathbf{A})$, use simulation to find the mean and standard deviation of h implied by various other sources of prior information, and favor parameter values that preserve the sign of h by adding $\log p(h)$ to the log prior for $\mathbf{A}$. We can also do the same for signs of impacts, selecting elements of $\tilde{\mathbf{H}}$ as additional h_j, or ratios of elements of $\tilde{\mathbf{H}}$ to incorporate prior information about magnitudes of plausible impacts as was done above.

3.4 Prior information about structural variances.

We follow Baumeister and Hamilton (2015) in using a natural conjugate form for the prior $p(\mathbf{D}|\mathbf{A})$, which turns out to be the product of independent inverse-gamma distributions,

$$p(\mathbf{D}|\mathbf{A}) = \textstyle\prod_{i=1}^{n} p(d_{ii}|\mathbf{A}) \tag{31}$$

$$p(d_{ii}^{-1}|\mathbf{A}) = \begin{cases} \frac{\tau_i(\mathbf{A})^{\kappa_i}}{\Gamma(\kappa_i)}(d_{ii}^{-1})^{\kappa_i - 1}\exp(-\tau_i(\mathbf{A})d_{ii}^{-1}) & \text{for } d_{ii}^{-1} \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

where d_{ii} denotes the row i, column i element of $\mathbf{D}$. The parameters κ_i and τ_i characterize the researcher's prior beliefs about structural variances, with κ_i/τ_i giving the analyst's expected value of d_{ii}^{-1} before seeing any data, while κ_i/τ_i^2 is the variance of this prior distribution. Small confidence in these prior beliefs would be represented by small values for κ_i and τ_i.

We set $\kappa_i = 2$, which gives our prior about the same influence as 4 observations of y_t and $\mathbf{x}_{t-1}$, and chose $\tau_i(\mathbf{A})$ to generate a value for $\tau_i(\mathbf{A})/\kappa_i$ equal to the variance of a univariate autoregression for $\mathbf{a}_i'\mathbf{y}_t$. Specifically, let $\hat{e}_{it}$ denote the residual of a fourth-order autoregression for series i and $\mathbf{S}$ the sample variance matrix of these univariate residuals ($s_{ij} = T^{-1}\sum_{t=1}^{T}\hat{e}_{it}\hat{e}_{jt}$). We set $\tau_i(\mathbf{A})$ equal to the ith diagonal element of $\kappa_i\mathbf{ASA}'$.

3.5 Prior information about lagged structural coefficients.

Prior beliefs about the lagged structural coefficients $\mathbf{B}$ are represented with conditional Gaussian distributions, $\mathbf{b}_i|\mathbf{A},\mathbf{D} \sim N(\mathbf{m}_i, d_{ii}\mathbf{M}_i)$:

$$p(\mathbf{B}|\mathbf{D},\mathbf{A}) = \textstyle\prod_{i=1}^{n} p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) \tag{32}$$

$$p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) = \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}} \exp[-(1/2)(\mathbf{b}_i - \mathbf{m}_i(\mathbf{A}))'(d_{ii}\mathbf{M}_i)^{-1}(\mathbf{b}_i - \mathbf{m}_i(\mathbf{A}))]. \tag{33}$$

Here $\mathbf{m}_i$ and $\mathbf{M}_i$ are parameters summarizing the researcher's prior information about the lagged coefficients in the ith structural equation. The vector $\mathbf{m}_i$ denotes our best guess before seeing the data as to the value of $\mathbf{b}_i$, where $\mathbf{b}_i'$ denotes row i of $\mathbf{B}$. The matrix $\mathbf{M}_i$ characterizes our confidence in these prior beliefs. A large variance would represent much uncertainty. Our values for $\mathbf{m}_i$ come from two different sources, the first being a "Minnesota prior" as in Doan, Litterman, and Sims (1984) and Sims and Zha (1999), and the second from specific information about the lagged coefficients in the monetary policy equation.

The Minnesota prior maintains that the single most useful variable for predicting $y_{i,t+1}$ is typically going to be the value of y_{it}. Insofar as some other variable y_{jt} also helps, its most recent value is likely to be more useful than its earlier values. Doan, Litterman and Sims suggested using random walks for the prior means, that is, a prior expectation that the reduced-form coefficient relating $y_{i,t+1}$ to y_{it} is likely to be unity. However, for our variables (the output gap, inflation, and interest rates) there is more of a tendency for mean reversion and so we instead use AR(1) processes with autoregressive coefficients $\phi = 0.75$. Specifically, our prior expectation is that elements of $\mathbf{b}_i$ after the first lag are likely to be 0 while the first 3 elements of $\mathbf{b}_i$ should be close to $\phi \mathbf{a}_i$.[20] We place increasing confidence in these prior beliefs for coefficients on higher-order lags, weighting our prior expectations for the first lag coefficients roughly equivalent to 5 observations and for the fourth lag coefficients equivalent to about 20 observations. We put practically no weight on prior information about the constant term (the last element of $\mathbf{b}_i$); for details see Appendix B.

We will also make use of direct prior knowledge about the lagged coefficients in the Taylor Rule (19), reflecting a belief that this equation should be similar to the popular specification (21). This would mean that the third element of $\mathbf{b}^m$ should equal ρ and all other elements of $\mathbf{b}^m$ (other than the last element associated with the constant term) are zero. That coefficients on $\mathbf{y}_{t-2}, \mathbf{y}_{t-3}, \mathbf{y}_{t-4}$ are zero is already implied by the Minnesota prior. That prior also had implications for the coefficients on $\mathbf{y}_{t-1}$ based on the expectation that each reduced-form equation might look like an AR(1) with autoregressive coefficient ϕ.[21] But equation (21) further implies that the coefficient on r_{t-1} should equal ρ. The weight of this prior is determined by the variance V_i in equations (34) and (35). We set $V_i = 0.1$, which gives this prior information a weight roughly equivalent to 3 observations; again see Appendix B for details. Using ρ in this way to inform estimation of the dynamic coefficients also helps identify the long-run Taylor parameters ψ^y and ψ^π.

---

[20] As in Sims and Zha (1998), note that if the ith structural equation took the form $\mathbf{a}_i'\mathbf{y}_t = \phi\mathbf{a}_i'\mathbf{y}_{t-1} + u_{it}$, then stacking the structural equations gives $\mathbf{A}\mathbf{y}_t = \phi\mathbf{A}\mathbf{y}_{t-1} + \mathbf{u}_t$. Recalling (3), we obtain the reduced form by premultiplying by $\mathbf{A}^{-1}$: $\mathbf{y}_t = \phi\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$.

[21] Specifically, these implied a prior expected value for the coefficient on y_{t-1} of $-\phi(1-\rho)\psi^y$, on π_{t-1} of $-\phi(1-\rho)\psi^\pi$, and on r_{t-1} of ϕ.

3.6 Impulse-response functions implied by the prior.

Table 2 summarizes implications of our prior for the structural impulse-response functions, with columns 1, 3, and 5 reporting the prior probability that each of the three structural shocks would increase each of the three variables for periods t, $t+1$, and $t+2$ following a shock in period t. Our prior places a very high probability that the effects have the expected signs on impact. But we have much less confidence that these effects persist into horizons $s = 1$ or 2. The red dashed lines in Figure 4 plot the median of our prior distribution for impulse-response functions through $s = 20$. Although the medians of our prior distribution for structural impulse-response functions die out fairly quickly, the uncertainty we associate with this prior information grows significantly as the horizon increases. For example, for the effect on inflation of a monetary shock, the width of a set around the median containing 90% of the prior probability is 39 basis points for $s = 0$, 115 basis points for $s = 4$, and 731 basis points for $s = 20$. Thus posterior inferences about the effects at longer horizons are almost all coming from the data and not the prior.

3.7 Empirical results.

Our analysis is based on quarterly data on $\mathbf{y}_t$ with the fourth-order VAR estimated over the period of the Great Moderation (t = 1986:Q1 to 2008:Q3). We used the algorithm in Baumeister and Hamilton (2015) to generate $N = 1$ million draws $\{\mathbf{A}^{(\ell)}, \mathbf{D}^{(\ell)}, \mathbf{B}^{(\ell)}\}_{\ell=1}^{N}$ from the posterior distribution $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$.

Posterior distributions for the 6 contemporaneous coefficients are plotted as histograms in Figure 2. The data turn out to be quite informative about the values of β^d, ψ^y, and ρ

but cause more modest revisions in our beliefs about other parameters.

Posterior impulse-response functions are plotted in Figure 4. The solid blue lines plot the median of the posterior distribution for any given horizon. Note that with informative priors, there is no ambiguity about reporting these solid lines as optimal point estimates despite the fact that the model is only set-identified. The shaded regions in Figure 4 represent 68% posterior credibility regions and the dashed lines indicate 95% regions.

The first column of Figure 4 summarizes the effects of a supply shock. This raises output and lowers inflation but has an unclear effect on interest rates. The data have been very informative about all three magnitudes, as can be seen by comparing the prior and posterior probabilities in columns 1 and 2 of Table 2. The second column of Figure 4 gives the effect of a demand shock, which raises output, inflation, and the interest rate. Effects on output and inflation of supply and demand shocks are quite persistent, with confidence about the signs of effects lasting well beyond one year. The third column in Figure 3 summarizes the effect of a one-unit increase in the monetary policy shock u_t^m on each of the three variables.[22]

These effects are small and do not seem to persist, and indeed the posterior median for the effect on output becomes positive after 1 year.

Figure 5 displays the historical decomposition of the output gap in terms of the contributions of the separate structural shocks. The dashed line is the observed value for the output

[22] Note that if there were no immediate effects of the policy on output or inflation, the fed funds rate would rise by 1% as a result of a monetary policy shock of one unit. However, our specification assumes that higher interest rates cause output and inflation to fall on impact, and these feed back into the interest rate. The Taylor Rule equation shifts up by 100 basis points, but within the quarter the economy moves along the new Taylor Rule equation with output falling 0.38% and inflation falling 0.17%, as a result of which in equilibrium the fed funds rate is only 67 basis points higher in the immediate response to the shock.

gap (in deviations from the sample mean). The solid line in the top panel is the posterior median contribution of supply shocks over the 10 years prior to the indicated date,[23] while the second and third panels give the contributions of demand and monetary policy shocks, respectively. The shaded regions and dashed lines denote 68% and 95% posterior credibility regions, respectively. To our knowledge, ours is the first paper to report such error bands in the very large literature using SVARs that are only set-identified. The high level of economic activity in the late 1980s is attributed primarily to strong demand, whereas the boom at the end of the 1990s is judged to be primarily driven by supply. Monetary policy seems to have typically played a minor role in output fluctuations.

Figures 6 and 7 report the decompositions for inflation and interest rates. Again the rising inflation of the late 1980s seems to have been driven by demand, while the low inflation of the late 1990s was primarily a supply-side development. The response of monetary policy to output and inflation as a result of exogenous shocks to demand, as opposed to deviations of the Fed from its traditional monetary policy rule, appear to be the primary cause of interest rate fluctuations.

We can summarize the average contribution of different shocks using variance decompositions. Table 3 reports the contribution of each of the three structural shocks to the mean-squared error of a one-year-ahead forecast of each of the three variables. Demand shocks account for 71% of the variance of interest rates and supply shocks account for about 2/3 of the variance of inflation. Demand shocks account for 60% of the variability of output

[23] That is, the panel plots the first element of (15) for $j = 1$ and $s = 40$.

and supply shocks another third. Monetary policy shocks are significantly less important for determining the paths of any of the three variables.

4 Sensitivity analysis.

Critics of the Bayesian approach sometimes question whether the prior information is "correct." A better way to formulate this concern is how reliable the prior information is, and this is directly controlled by the parameters that represent the prior information.

The blue line in Figure 8 plots the Student t distribution ($\mu = 0.75, \sigma = 0.4, \nu = 3$) that we used to represent prior information about β^d. If we regarded this information as less reliable, we would use a bigger value of σ. For $\sigma = 10$, the prior information is modeled as completely unreliable and the prior for β^d would have no influence on the posterior inference. As $\sigma \to 0$, the prior information is treated as perfectly reliable. If we set $\mu_\alpha = \mu_\gamma = 0$, $\sigma_\alpha = \sigma_\gamma = 0$, and $\sigma_\beta = \sigma_{\psi^y} = \sigma_{\psi^\pi} = 10$ we would obtain the traditional Cholesky identification as a special case of the general approach followed here.[24] The concern that prior information may not be correct is not a criticism of Bayesian methods but instead is a criticism of the traditional identifying assumption that $\sigma = 0$. Indeed, it is precisely because prior information is not perfectly reliable that more researchers should be using Bayesian methods!

Of course, if no prior information is reliable we would be back in the position of being unable to say anything about the effects of policy. But since our application draws on a little

[24] In Baumeister and Hamilton (2017) we demonstrated this by numerically replicating an influential analysis of the economic determinants of oil price fluctuations that had used a Cholesky identification.

bit of information about a large number of elements, we can investigate how the inference would change if we completely wipe out the influence of any one element of the prior.

The upper left panel of Figure 9 shows the estimated effect of a monetary policy shock on output if we replaced $\sigma_\alpha = 0.4$ in the baseline specification (shown in red for comparison) with $\sigma_\alpha = 10$ (shown in blue), holding all other elements of the prior fixed. We would draw essentially the same conclusion about the effects of monetary policy. Figure D1 in online Appendix D shows the way this change would affect our inference about the effects of all the shocks on all the variables. The upper right panel of Figure 9 returns to $\sigma_\alpha = 0.4$ but now takes $\sigma_\beta = 10$. The next three panels throw out the contribution of the priors about γ^d, ψ^y, and ψ^π, respectively. For ρ we replace the Beta$(2.6, 2.6)$ prior with a uniform prior over $(0, 1)$. The last two panels set $\zeta_{h_1} = 0$ or $\zeta_{h_2} = 0$, respectively. No single element of the prior has any material influence on our conclusions about the effects of monetary policy, though of course collectively the prior information played a critical role in our ability to draw structural interpretations from the correlations in the data.

5 Conclusion.

Structural inference is only possible if we have prior information about the underlying economic model and mechanisms. The traditional approach to identification acts as though this prior information enables us to know some features of the structure with certainty. In this paper we have proposed generalizing this approach to acknowledge doubts about the prior information. In making this generalization, the model becomes only set-identified.

But we can still form an inference based on what we do know and incorporate uncertainty about the model itself into any statistical conclusions. In this paper we investigated statistical inference about impulse-response functions, historical decompositions, and variance decompositions in such a setting using Bayesian statistical decision theory, and showed that for reasonable loss functions these can be estimated pointwise from the Bayesian posterior mean or median of the relevant magnitudes. We noted that this is implicitly what has been done by hundreds of researchers using sign-restricted VARs, but argued that the methods only make sense when the prior is explicit rather than implicit. We illustrated these methods using a simple macroeconomic model, and concluded that monetary policy shocks played a relatively minor role in influencing output and inflation during the period of the Great Moderation.

References

Amir-Ahmadi, Pooyan, and Harald Uhlig (2015). "Sign Restrictions in Bayesian FAVARs with an Application to Monetary Policy Shocks," NBER Working Paper 21738.

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (forthoming). "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," *Econometrica*.

Azzalini, Adelchi, and Antonella Capitanio (2003). "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution," *Journal of the Royal Statistical Society, Series B* (*Statistical Methodology*) 65: 367-389.

Baumeister, Christiane, and James D. Hamilton (2015). "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica* 83: 1963-1999.

Baumeister, Christiane, and James D. Hamilton (2017). "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," NBER working paper 24167.

Benati, Luca (2008). "Investigating Inflation Persistence Across Monetary Regimes," *Quarterly Journal of Economics* 123: 1005-1060.

Benati, Luca, and Paolo Surico (2009). "VAR Analysis and the Great Moderation," *American Economic Review* 99: 1636-1652.

Canova, Fabio, and Gianni De Nicoló (2002). "Monetary Disturbances Matter for Business Fluctuations in the G-7," *Journal of Monetary Economics* 49: 1131-1159.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy* 113: 1-45.

Del Negro, Marco, and Frank Schorfheide (2004). "Priors from General Equilibrium Models for VARs," *International Economic Review* 45(2): 643-673.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1984). "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews* 3: 1-100.

Drèze, Jacques H. (1974). "Bayesian Theory of Identification in Simultaneous Equations Models," in: *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage,* pp. 159-174, edited by Stephen E. Fienberg and Arnold Zellner. Amsterdam: North-Holland.

Drèze, Jacques H., and Juan-Antonio Morales (1976). "Bayesian Full Information Analysis of Simultaneous Equations," *Journal of the American Statistical Association* 71: 919-923.

Fry, Renée, and Adrian Pagan (2011). "Sign Restrictions in Structural Vector Autoregressions: A Critical Review," *Journal of Economic Literature* 49(4): 938-960.

Gafarov, Bulat, Matthias Meier, and José Luis Montiel Olea (2016). "Projection Inference for Set-Identifed SVARs," working paper, New York University.

Gafarov, Bulat, Matthias Meier, and José Luis Montiel Olea (forthcoming). "Delta-Method Inference for a Class of Set-Identified SVARs," Journal of Econometrics.

Giacomini, Raffaella, and Toru Kitagawa (2015). "Robust Inference about Partially Identified SVARs," working paper, University College London.

Giordani, Paolo (2004). "An Alternative Explanation of the Price Puzzle," *Journal of Monetary Economics* 51: 1271–1296.

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

Inoue, Atsushi, and Lutz Kilian (2013). "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics* 177: 1-13.

Jordá, Oscar (2009). "Simultaneous Confidence Regions for Impulse Responses," *Review of Economics and Statistics* 91: 629–647.

Kadane, Joseph B. (1974). "The Role of Identification in Bayesian Theory," in: *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage,* pp. 175-191, edited by Stephen E. Fienberg and Arnold Zellner. Amsterdam: North-Holland.

Kline, Brendan, and Elie Tamer (2016). "Bayesian inference in a class of partially identified models," *Quantitative Economics* 7: 329-366.

Lubik, Thomas, and Frank Schorfheide (2004). "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review* 94(1): 190-217.

Lütkepohl, Helmut (1990). "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics* 72(1): 116-25.

Lütkepohl, Helmut (2005). *New Introduction to Multiple Time Series Analysis,* Berlin: Springer.

Montiel Olea, Jose L., and Mikkel Plagborg-Møller (2017). "Simultaneous Confidence Bands: Theoretical Comparisons and Suggestions for Practice," working paper, Princeton University.

Moon, Hyungsik Roger, Frank Schorfheide, and Eleonora Granziera (forthcoming). "Inference for VARs Identified with Sign Restrictions," *Quantitative Economics.*

Poirier, Dale J. (1998). "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14: 483-509.

Rotemberg, Julio J., and Michael Woodford (1997). "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in: *NBER Macroeconomics Annual*, Volume 12, pp. 297-344, edited by Ben S. Bernanke and Julio J. Rotemberg. Cambridge, MA: M.I.T. Press.

Rubio-Ramírez, Juan, Daniel F. Waggoner, and Tao Zha (2010). "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies* 77(2): 665-696.

Sims, Christopher A., and Tao Zha (1998). "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review* 39: 949-968.

Sims, Christopher A., and Tao Zha (1999). "Error Bands for Impulse Responses," *Econometrica* 67: 1113-1155.

Song, Kyungchul (2014). "Point Decisions for Interval-Identified Parameters," *Econometric Theory* 30: 334-356.

Taylor, John B. (1993). "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39: 195-214.

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.

Table 1. Priors for contemporaneous coefficients.

Parameter	Meaning	Prior mode	Prior scale	Sign restriction
		Student t distribution with 3 degrees of freedom		
α^s	Effect of π on supply	2	0.4	$\alpha^s \geq 0$
β^d	Effect of π on demand	0.75	0.4	none
γ^d	Effect of r on demand	-1	0.4	$\gamma^d \leq 0$
ψ^y	Fed response to y	0.5	0.4	$\psi^y \geq 0$
ψ^π	Fed response to π	1.5	0.4	$\psi^\pi \geq 0$
		Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$		
ρ	Interest rate smoothing	0.5	0.2	$0 \leq \rho \leq 1$

Table 2. Prior and posterior probabilities that the impact of a specified structural shock on the indicated variable is positive at horizons s = 0, 1, and 2.

	Supply shock		*Demand shock*		*Monetary policy shock*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Prior	Posterior	Prior	Posterior	Prior	Posterior
Variable						
			***s* = 0**			
y	0.851	1.000	1.000	1.000	0.000	0.000
π	0.000	0.000	1.000	1.000	0.000	0.000
r	0.008	0.229	1.000	1.000	0.999	1.000
			***s* = 1**			
y	0.717	1.000	0.994	1.000	0.037	0.079
π	0.006	0.000	0.961	1.000	0.117	0.046
r	0.054	0.374	0.965	1.000	0.981	1.000
			***s* = 2**			
y	0.617	1.000	0.974	1.000	0.143	0.206
π	0.021	0.000	0.879	1.000	0.272	0.078
r	0.156	0.478	0.869	1.000	0.916	1.000

Table 3. Decomposition of variance of 4-quarter-ahead forecast errors.

	Supply	Demand	Monetary policy
Output gap	**0.36** [35%] (0.10, 0.84)	**0.62** [60%] (0.34, 1.10)	**0.05** [5%] (0.01, 0.19)
Inflation	**0.38** [69%] (0.20, 0.68)	**0.16** [28%] (0.05, 0.36)	**0.02** [3%] (0.00, 0.09)
Fed funds rate	**0.02** [1%] (0.00, 0.16)	**0.94** [71%] (0.37, 1.74)	**0.37** [28%] (0.11, 0.92)

Notes. Estimated contribution of each structural shock to the 4-quarter-ahead median squared forecast error of each variable in bold, and expressed as a percent of total MSE in brackets. Parentheses indicate 95% credibility intervals.

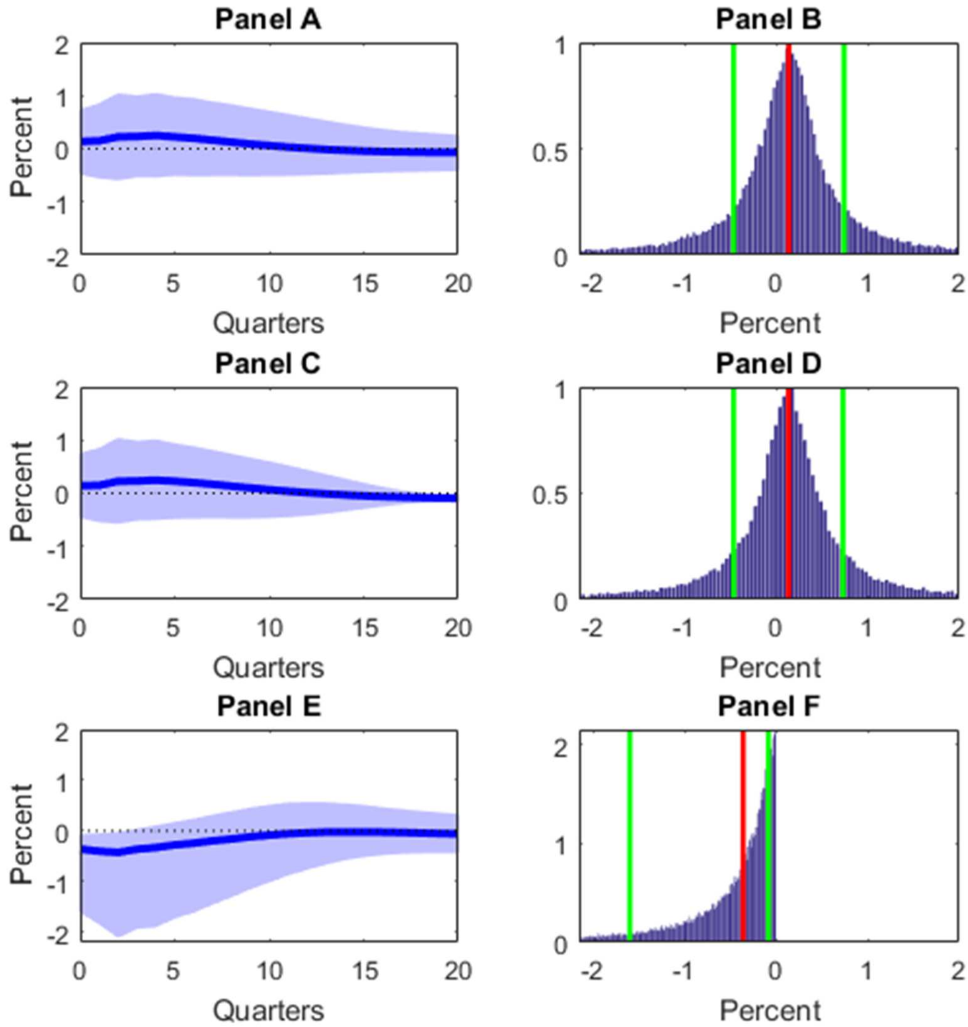

Figure 1. Estimating the effects of monetary policy ostensibly without making any assumptions. Panel A: Response of output gap to a 25-basis-point monetary contraction based on median and 68% of generated draws with reduced-form parameters $\mathbf{\Omega}$ and $\mathbf{\Phi}$ drawn from Normal-inverse-Wishart posterior but imposing no sign restrictions at all. Panel B: Histogram (in blue), median (in red), and 16% and 84% quantiles (in green) of response at horizon $s = 0$ from Panel A. Panels C and D: Same as panels A and B but with $\mathbf{\Omega}$ and $\mathbf{\Phi}$ fixed at maximum likelihood estimates. Panels E and F: Response of output gap using only the sign restriction that a monetary contraction lowers the output gap on impact.

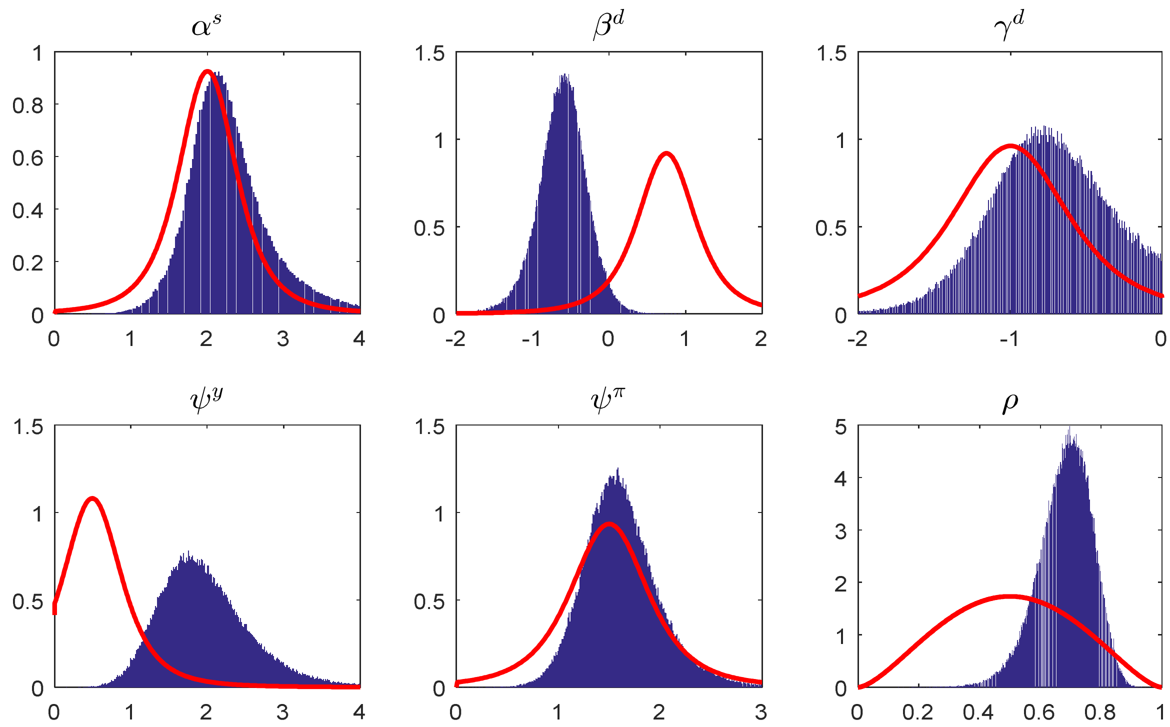

Figure 2. Prior distributions (red lines) and posterior distributions (blue histogram) for contemporaneous coefficients.

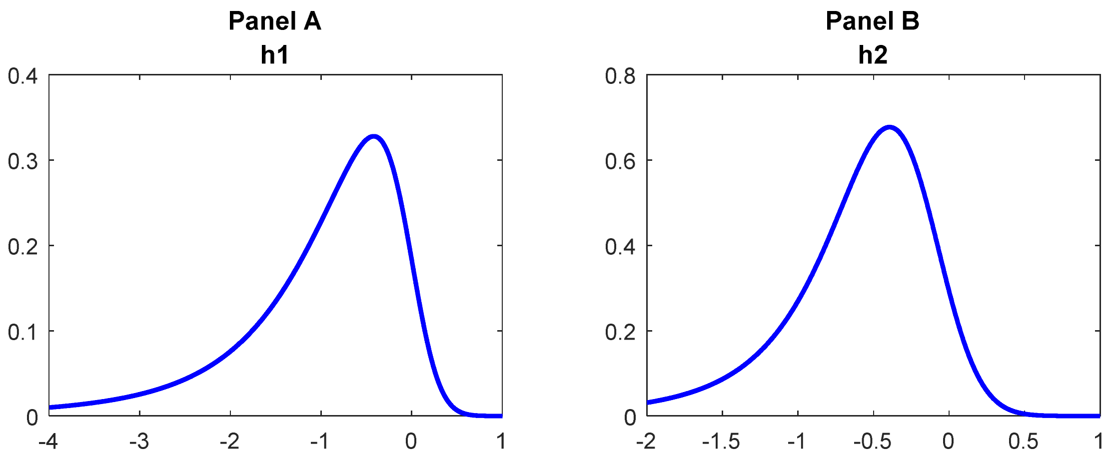

Figure 3. Asymmetric *t* distributions representing priors for impact coefficients. Panel A: Prior for (1,1) element of adjoint of **A** (governs the equilibrium response of output to a favorable supply shock). Plots the density in equation (28) for $\mu_h = -0.1, \sigma_h = 1, \nu_h = 3, \lambda_h = -4$. Panel B: Prior for ratio of (1,3) to (3,3) elements of adjoint of **A** (governs the size of equilibrium response of output to monetary contraction). Plots the density in equation (28) for $\mu_h = -0.3, \sigma_h = 0.5, \nu_h = 3, \lambda_h = -2$.

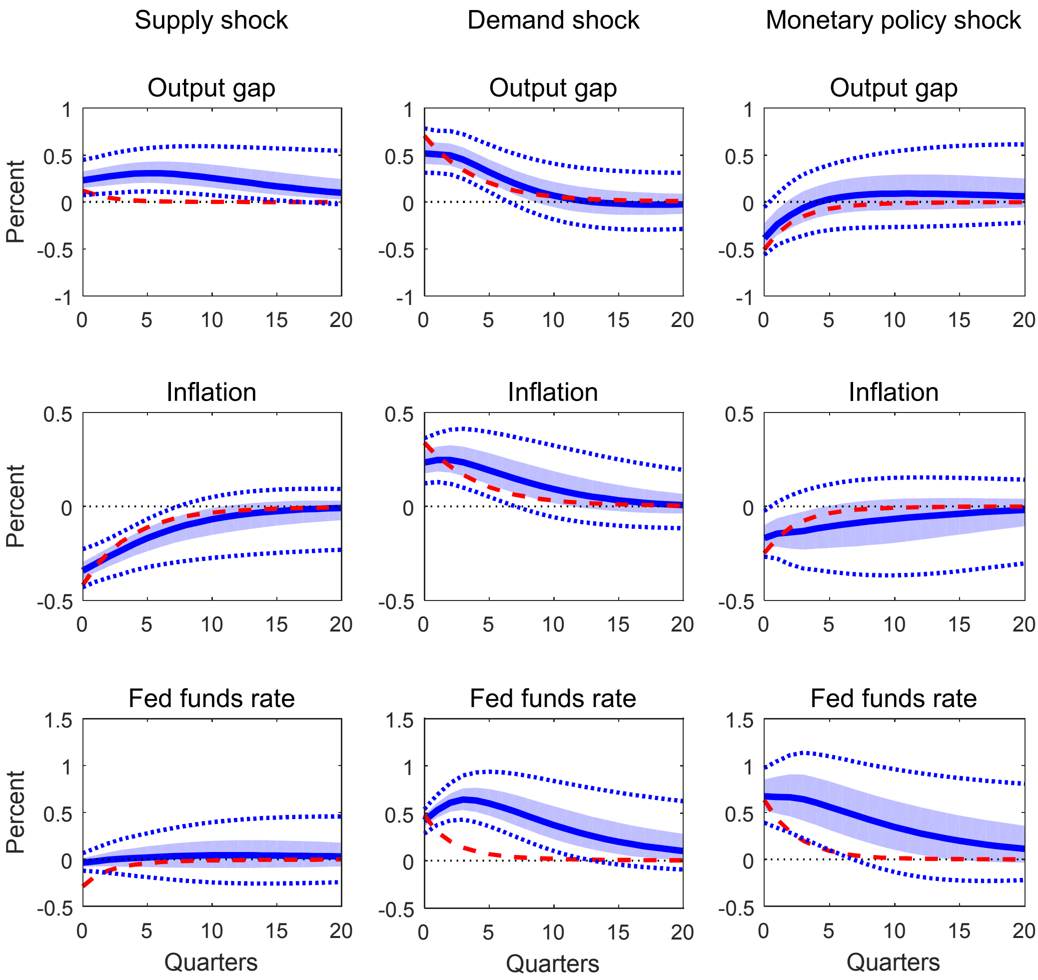

Figure 4. Structural impulse-response functions for 3-variable VAR. Solid blue lines: posterior median. Shaded regions: 68% posterior credibility set. Dotted blue lines: 95% posterior credibility set. Dashed red lines: prior median.

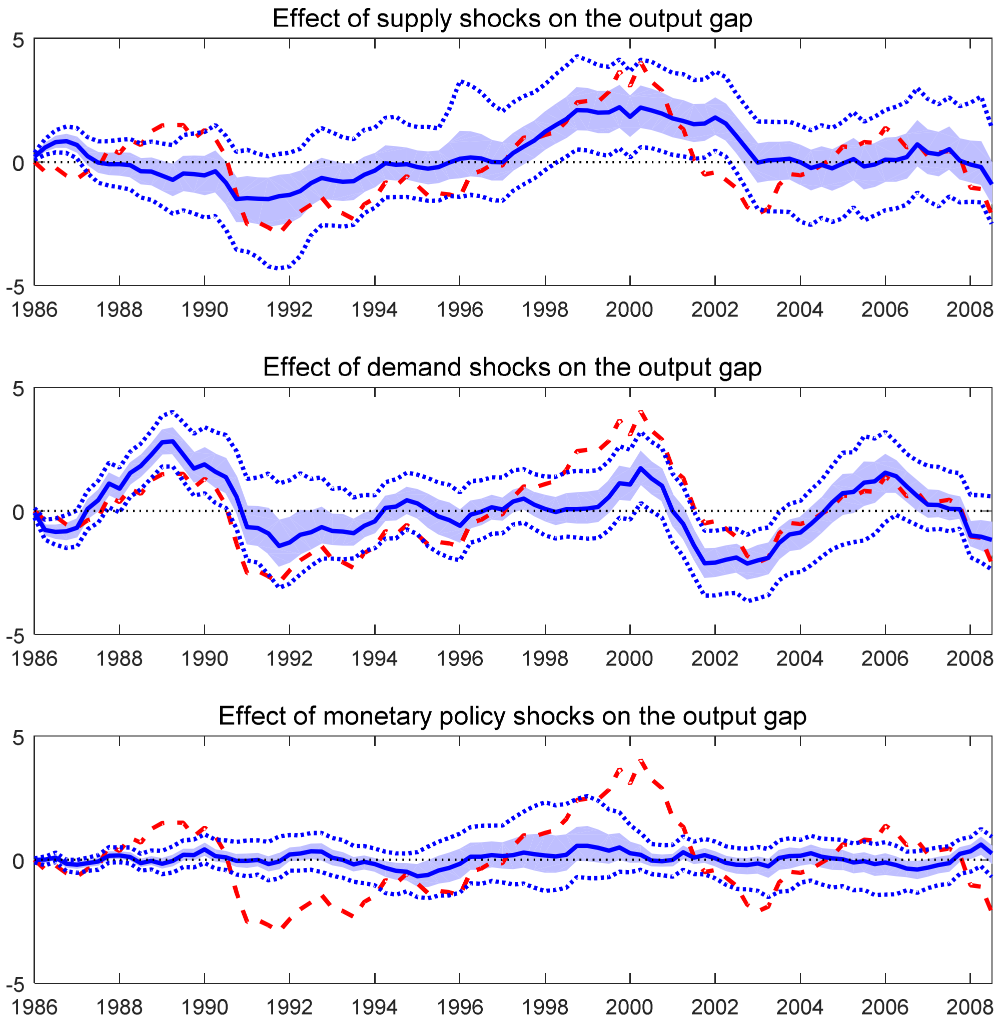

Figure 5. Portion of historical variation in output gap attributed to each of the structural shocks. Dashed red: actual value for the deviation of output gap from its mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

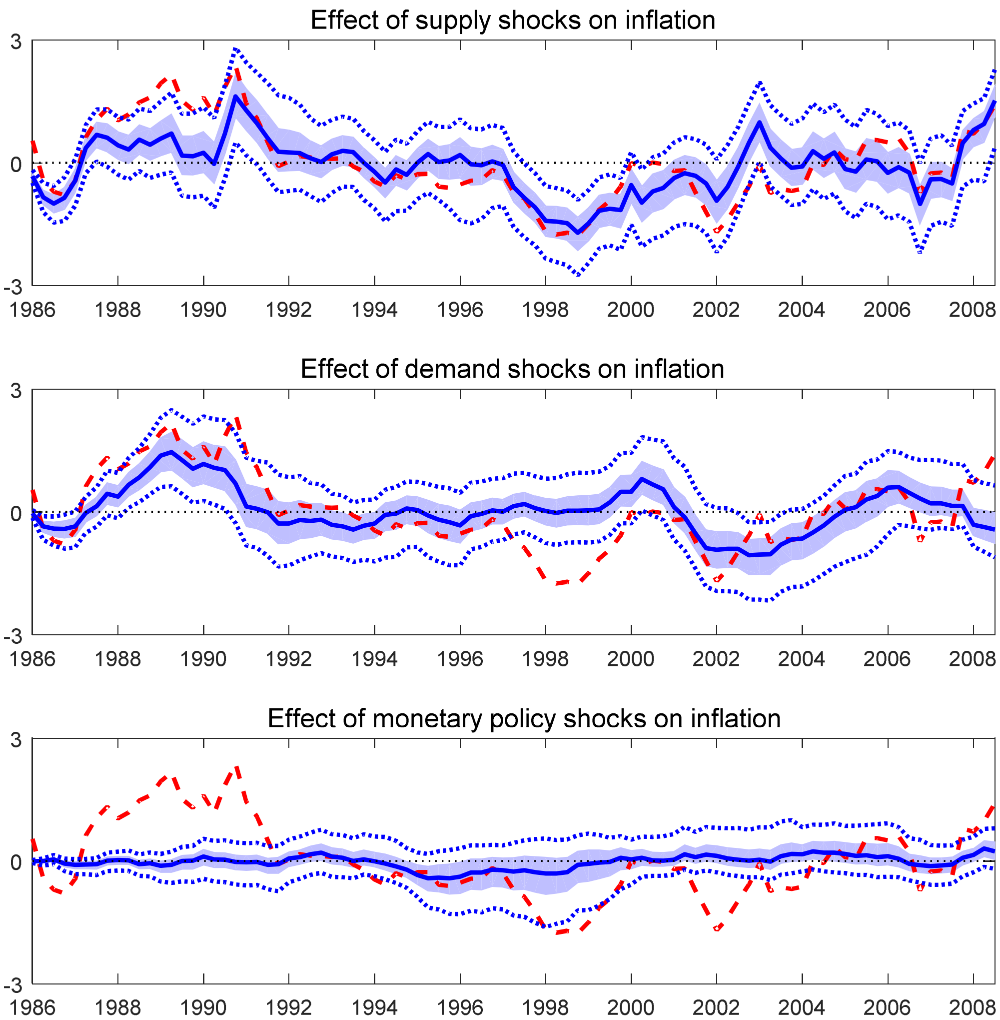

Figure 6. Portion of historical variation in inflation attributed to each of the structural shocks. Dashed red: actual value for the deviation of inflation from its mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

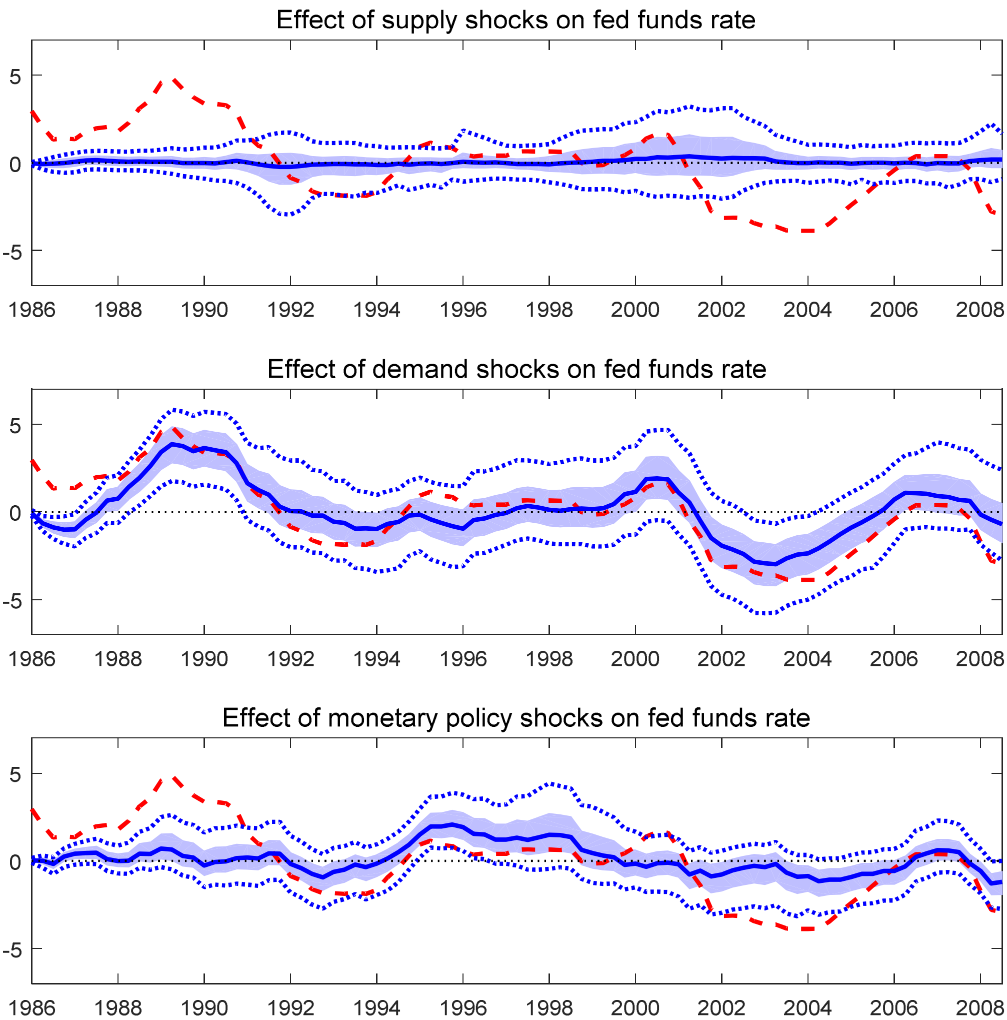

Figure 7. Portion of historical variation in fed funds rate attributed to each of the structural shocks. Dashed red: actual value for the deviation of fed funds rate from its mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

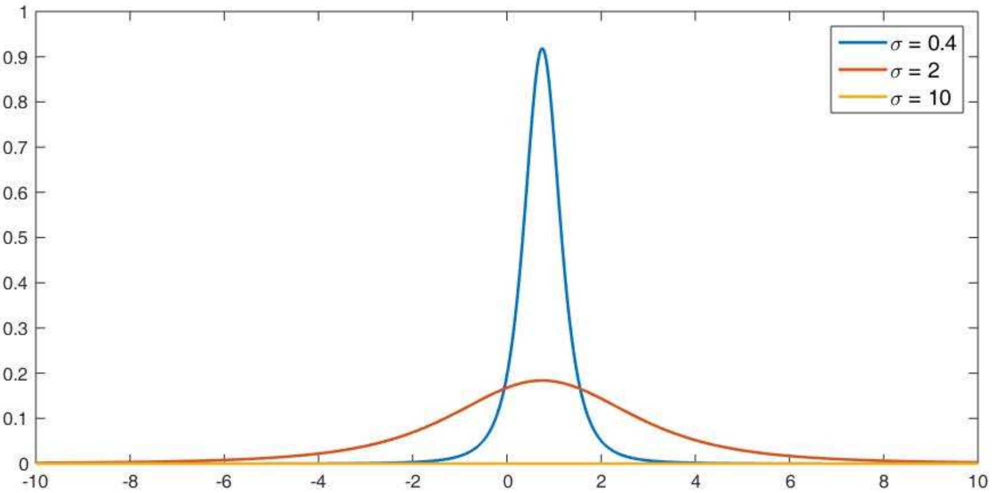

Figure 8. Plot of Student *t* density with location parameter 0.75, 3 degrees of freedom, and scale parameter of 0.4, 2, or 10.

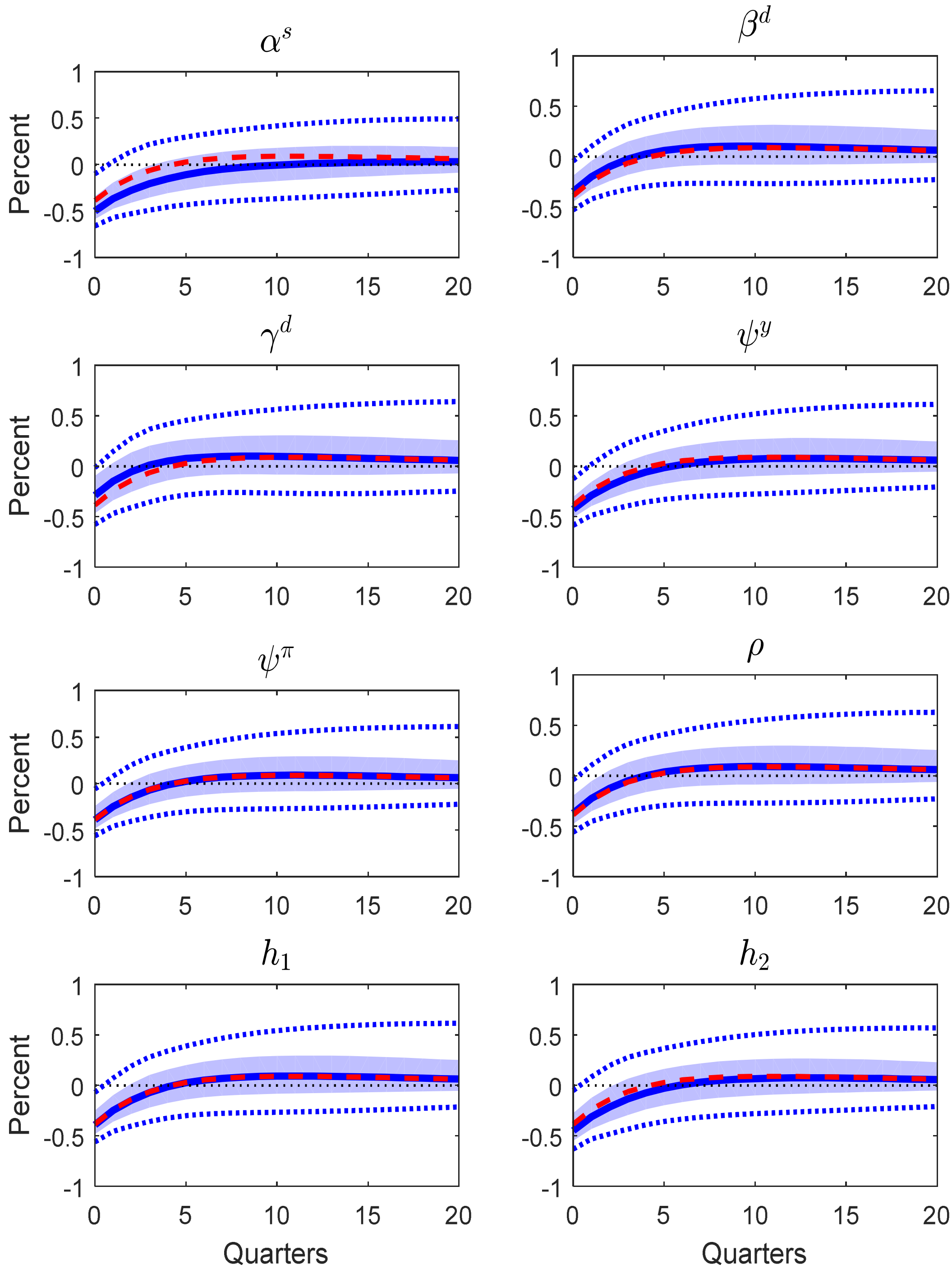

Figure 9. Sensitivity analysis. Response of the output gap to a contractionary monetary policy shock with an uninformative prior about the parameter indicated in the label.
Solid blue lines: posterior median. Shaded regions: 68% posterior credibility set. Dotted blue lines: 95% posterior credibility set. Dashed red lines: Posterior median from benchmark specification.

Appendix A. Traditional sign-restriction algorithm.

Here we describe the sign-restriction algorithm developed by Rubio-Ramírez, Waggoner, and Zha (2010) that was used to generate the impulse responses and histograms of the impact effect of a contractionary monetary policy shock in Figure 1.

Let $\mathbf{K}$ denote an $n \times n$ matrix whose elements are random draws from independent standard Normal distributions. Take the QR decomposition of $\mathbf{K}$ such that $\mathbf{K} = \mathbf{QR}$ where $\mathbf{R}$ is an upper triangular matrix whose diagonal elements have been normalized to be positive and $\mathbf{Q}$ is an orthonormal matrix ($\mathbf{QQ}'= \mathbf{I}_n$). Let $\mathbf{P}$ be the Cholesky factor of the reduced-form variance-covariance matrix $\mathbf{\Omega}$ (so that $\mathbf{\Omega} = \mathbf{PP}'$) and generate a candidate impact matrix $\mathbf{H} = \mathbf{PQ}$.

In the absence of any sign restrictions, keep every draw of $\mathbf{H}$ and compute impulse responses $\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{u}_t'} = \mathbf{\Psi}_s \mathbf{H} \equiv \mathbf{H}_s$ for $\mathbf{\Psi}_s$ the matrix in equation (7). Absent any identifying assumptions, the impulse responses of variable y_i after any one-standard-deviation structural shock u_j are the same (see Baumeister and Hamilton, 2015, equation (33)), so that it does not matter which column of $\mathbf{H}_s$ is selected for the monetary policy shock. To obtain the dynamic effect of a 25 basis point increase in the federal funds rate on the output gap, divide the entire impulse response of the output gap by the contemporaneous response of the fed funds rate, and scale this response by multiplying it by 0.25; then sort the draws and compute the median and the 16^{th} and 84^{th} percentiles.

To impose only the sign restrictions that a monetary policy shock moves the output gap and the federal funds rate in opposite direction, we keep the matrix $\mathbf{H}$ if the (1,1) and (3,1) elements of $\mathbf{H}$ are of opposite sign, and throw out $\mathbf{H}$ and draw a new matrix if they are of the same sign. For the accepted draws we compute the impulse responses in the same way as described above.

Following Uhlig (2005, p. 410), we account for estimation uncertainty of the reduced-form VAR parameters by taking draws for $(\mathbf{\Phi}, \mathbf{\Omega})$ from a Normal-inverse Wishart posterior and apply the QR algorithm to each reduced-form posterior draw until we have $50,000$ accepted draws.

Appendix B. Details of implementing priors on lagged structural coefficients.

Baumeister and Hamilton (2015) showed that the mean $\mathbf{m}_i^*$ and variance $d_{ii}\mathbf{M}_i^*$ of the posterior distribution $p(\mathbf{b}_i|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T)$ for the lagged coefficients of the ith structural equation can be found from an OLS regression of $\tilde{\mathbf{Y}}_i$ in their equation (47) on $\tilde{\mathbf{X}}_i$ in equation (48). For the current application these take the form

$$\underset{[(T+13+1)\times 1]}{\tilde{\mathbf{Y}}_i} = \begin{bmatrix} \mathbf{a}_i'\mathbf{y}_1 & \cdots & \mathbf{a}_i'\mathbf{y}_T & \mathbf{m}_i'\mathbf{P}_i & r_i/\sqrt{V_i} \end{bmatrix}' \tag{34}$$

$$\underset{[(T+13+1)\times 13]}{\tilde{\mathbf{X}}_i} = \begin{bmatrix} \mathbf{x}_0 & \cdots & \mathbf{x}_{T-1} & \mathbf{P}_i & \mathbf{e}_i/\sqrt{V_i} \end{bmatrix}' \tag{35}$$

where $\mathbf{a}_i'$ denotes the ith row of $\mathbf{A}$ in (24). Prior information about lagged structural coefficients comes from two sources. Information about the reduced form gives us an expectation that $\mathbf{b}_i$ could be similar to $\mathbf{m}_i = 0.75\boldsymbol{\eta}'\mathbf{a}_i$ where

$$\underset{(3\times 13)}{\boldsymbol{\eta}} = \begin{bmatrix} \underset{(3\times 3)}{\mathbf{I}_3} & \underset{(3\times 10)}{\mathbf{0}} \end{bmatrix}.$$

Our confidence in this prior information about the reduced form is captured by $\mathbf{P}_i$, which we specified as a diagonal matrix whose value associated with the coefficient on the ℓth lag of variable j is $\ell^{\lambda_1}s_{jj}/\lambda_0$ where s_{jj} is the estimated innovation standard deviation of a univariate fourth-order autoregression fit to variable j. We set λ_0, the parameter controlling the overall tightness of the prior, equal to 0.1, and set λ_1, which governs how quickly the prior for lagged coefficients tightens to zero as the lag ℓ increases, equal to unity. The last diagonal element of $\mathbf{P}_i$, which is the reciprocal of the standard deviation of the prior for the intercept in the ith structural equation, is taken to be $1/(\lambda_0\lambda_3)$, where we set $\lambda_3 = 100$.

In addition we have direct beliefs about the lagged structural coefficients as captured by the terms r_i and V_i in equations (34) and (35). This added information is used only for $i = 3$, the monetary policy rule, where the expectation is that the third element of $\mathbf{b}_3$ should be close to ρ. This is implemented by taking $\mathbf{e}_i$ in equation (35) to be column 3 of $\mathbf{I}_{13}$ and r_i in equation (34) equal to ρ. Our confidence in this prior information is captured by the value of V_i, with a smaller value for V_i representing greater confidence in the prior information.