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 March 30, 2017

Abstract

We demonstrate that formal Bayesian analysis of structural vector autoregressions offers clear guidance for inference about impulse-response functions and historical decompositions even if the model is not identified in the classical sense. Bayesian posterior probabilities describe uncertainty coming not just from randomness in the data but also uncertainty about the model itself. We illustrate these methods with 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

JEL Classifications: C32, E52, C11

*We thank Marek Jarocinski 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, 722 Flanner Hall, Notre Dame, IN 46556, cjs-baumeister@gmail.com

[‡]Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA, jhamilton@ucsd.edu

[§]Corresponding author, Tel: +1 858.534.5986

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):

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

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

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

The parameters of the VAR, $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t)$, 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 $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ 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 $\boldsymbol{\Omega}$.

Identification requires drawing on additional information about the structural model. For example, $\boldsymbol{\theta}$ would be identified if we knew that $\mathbf{D}$ is diagonal and $\mathbf{A}$ is lower triangular, 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 using less than a complete set of identifying assumptions. Canova and De Nicoló (2002), Uhlig (2005), Rubio-Ramírez, Waggoner and Zha (2010) and Arias, Rubio-Ramírez and Waggoner (2016) have developed algorithms for inference using in part restrictions on only the signs of certain objects of interest. This approach, which we will refer to as a sign-restricted structural vector autoregression, or SR-VAR has now been used in hundreds of empirical studies. Such an incomplete identification scheme means that there is a set of values for the answer to many structural questions of interest all of which would be equally consistent with the observed data. Even with an infinite sample of observations, we could only know that $\boldsymbol{\theta}$ falls within some particular set (referred to as the "identified set") but we would not know the value of $\boldsymbol{\theta}$ itself.

One option is just to report estimates of the identified set itself. Moon et al. (2011) and Gafarov, Meier and Olea (2016a) developed algorithms to implement this from a frequentist perspective, 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 Olea (2016b) suggested reporting the Bayesian posterior inference across a set of alternative possible Bayesian prior distributions as a robust Bayesian approach to posterior inference about the identified set.

Nevertheless, hundreds of applied studies using SRVARs have chosen to report point estimates and not just sets. It is unclear what the theoretical justification for this practice could be. Song (2014) noted that if we have a minimax loss function, it could be reasonable to report the midpoint of the identified set, though in many applications of SRVARs the identified set is unbounded. Fry and Pagan (2011) raised the concern that the point estimates that users of SRVARs typically report for impulse-response functions are not in fact consistent with any fixed value for $\boldsymbol{\theta}$. Inoue and Kilian (2013) proposed reporting the mode of the posterior density of the impulse-response function as one way to address the Fry and Pagan critique.

In this paper we suggest an alternative approach, which is to regard a proper Bayesian prior distribution for $\boldsymbol{\theta}$ as a strict generalization of both conventional identified and sign-restricted structural VARs. For example, one can view Cholesky identification as a special case of a prior distribution on $\boldsymbol{\theta}$, namely a dogmatic prior which maintains that there is no possibility that any element of $\mathbf{A}$ above the principal diagonal is nonzero. A generalization of this approach would acknowledge some reasonable doubts about this identifying assumption and replace it with a prior that regards these elements as likely close, but not necessarily exactly equal, to zero. And instead of a simple SRVAR, we could use for example not only the prior information that supply elasticities cannot be negative, but further treat extremely large positive values as relatively unlikely.

It has long been understood (Kadane, 1974; Poirier, 1998) that if one applies Bayesian methods to a model that is unidentified, the data will be uninformative about certain questions. For such questions, the Bayesian posterior inference will be the same as the prior, even with an infinite sample of data. Nevertheless, if the prior represents an accurate representation of our understanding of the economic structure, for example, of the confidence (strong or weak) that we place in conventional identifying assumptions, then statistical decision theory gives the optimal way to incorporate that prior information about the model.[2] In essence what we should be looking at is a weighted average of the inferences that would emerge from different observationally equivalent values for $\boldsymbol{\theta}$ with smaller weights on those values that we judge a priori to be less plausible (for example, zero weights on negative supply elasticities and very low weights on extremely large positive supply elasticities). This principle gives a basis for choosing a sensible point estimate to report within the identified set. Furthermore, calculation of Bayesian posterior probabilities allows us to report not just uncertainty that comes from randomness of the data but also uncertainty about the model itself.

In Section 2 we demonstrate that for typical loss functions, the optimal estimate of a structural impulse-response function in an unidentified model can be obtained from the Bayesian posterior mean or posterior median, calculated pointwise at each horizon. This provides a formal justification for the procedure typically adopted by users of SRVARs, but with an important caveat. The procedure is only appropriate if the researcher is willing to defend a particular prior distribution about $\boldsymbol{\theta}$, specifying exactly what the prior is informative about and what it is not. Using the posterior mean or median cannot be justified if the prior

[2] For an early discussion of some of these insights see Drèze (1974).

on $\boldsymbol{\theta}$ is simply a mechanical artifact of the algorithm used to implement sign restrictions, as shown by Baumeister and Hamilton (2015). Our results further demonstrate why the Fry and Pagan (2011) concern is misplaced, and contribute to the discussions by Sims and Zha (1999), Lütkepohl (1990, 2005) and Jordà (2009) on how to estimate and report uncertainty about impulse-response functions.

In Section 3 we show that analogous results hold for calculating the contributions of individual structural shocks to a given historical episode of interest. To our knowledge, every previous application of SRVARs that reported historical decompositions gave point estimates without error bands, despite the fact that, given the assumptions espoused by the users of this method, intervals are the only objects the researcher could actually defend reporting. 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 confidence in this estimate as long as the prior used in the analysis is explicitly defended. Section 4 derives parallel results for variance decompositions.

Section 5 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 calibrations 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 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 year, and monetary policy shocks typically make only a modest contribution to economic fluctuations.

2 Inference and credibility sets for impulse-response functions.

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 $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{D})$,

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

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 5.

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

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

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 $\mathbf{A}^{-1}\mathbf{B}$.[3]

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

[3] $\boldsymbol{\Psi}_s$ 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 $\mathbf{A}^{-1}\mathbf{B}$.

associated with our choice of $\hat{\mathbf{h}}_{ij}$ where this expectation is taken with respect to the posterior distribution of $\boldsymbol{\theta}$:

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

Here $g(\mathbf{h}_{ij}, \hat{\mathbf{h}}_{ij})$ is a loss function summarizing how upset we would be if our estimate of the 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[4] 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

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

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

Alternatively, if our loss function is instead

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

[4] Notice that

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

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

for any set of positive weights $\{\omega_s\}_{s=0}^{S-1}$, it is not hard to show[5] that element s of the optimal estimate $\hat{\mathbf{h}}_{ij}$ is the posterior median of $h_{ij}^s(\boldsymbol{\theta}^{(\ell)})$.[6]

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

[5] For this case we have

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

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

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

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 of impulse-response coefficients $\mathbf{h}$.

3 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])

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

for $\boldsymbol{\Psi}_s(\boldsymbol{\theta})$ the nonorthogonalized impulse-response matrix in (7). 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

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

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.

4 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

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

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

for $\mathbf{h}_j(s;\boldsymbol{\theta})$ the jth column of $\mathbf{H}_s(\boldsymbol{\theta})$ 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.

5 Bayesian inference in a 3-variable macro model.

Here we illustrate these methods using a commonly studied three-variable macroeconomic model.[7] 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).

5.1 Model description.

The system consists of a Phillips Curve,

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

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.

[7] 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).

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. For example, Cholesky identification can be represented as a special case of a Bayesian prior in which we know with certainty that α^s and γ^d are zero but have no information at all about β^d, ζ^y, or ζ^π.[8] 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 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 $p(\mathbf{B}|\mathbf{A},\mathbf{D})$ respectively in the following three subsections.

[8] In Baumeister and Hamilton (2016) we showed in an analysis of the economic effects of oil price shocks that applying our Bayesian algorithm for such a specification of the prior reproduces exactly the OLS Cholesky results that would be calculated using the standard frequentist algorithm.

5.2 Priors for contemporaneous 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 3 degrees of freedom, truncated to be positive. This density is plotted as a red curve in the lower-left panel of Figure 1. This 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. The low degrees of freedom give the density a very fat tail, so that even quite large values of ψ^y are not ruled out. 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 1. For the smoothing parameter ρ we use a beta(13,5) distribution (right bottom panel). This has a mode of 0.75, standard deviation around 0.1, and forces ρ to be between 0 and 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 above.

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 5.4. 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

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

where 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.[9] One option would be to take a completely specified dynamic stochastic general 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 $\pi_{t+1|t} = c^\pi + \phi^\pi \pi_t$,

[9] See for example equation (1) in Lubik and Schorfheide (2004).

where our prior expectation is $\phi^y = \phi^\pi = \phi = 0.75$. Substituting these expressions into (23) gives

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

where $\mu^d = c^d + c^y + \tau c^\pi$ and $\tilde{z} = z/(1-\phi)$. Many macro models assume an intertemporal elasticity of substitution of $\tau = 0.5$, which together with a value of $\phi = 0.75$ would imply $\tilde{\tau} = 2$. This leads us to use a Student t prior for γ^d in (18) with mode -2, 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 1.5. 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. In summary, for

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

our prior is given by

$$p(\mathbf{A}) = p(\alpha^s)p(\beta^d)p(\gamma^d)p(\psi^y)p(\psi^\pi)p(\rho).$$

The individual priors are summarized in Table 1 and plotted as red curves in Figure 1.

5.3 Priors for 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,

$$p(\mathbf{D}|\mathbf{A}) = \textstyle\prod_{i=1}^{n} p(d_{ii}|\mathbf{A}) \tag{25}$$

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

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 $\kappa_i\mathbf{ASA}'$.

5.4 Priors for lagged structural coefficients.

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

$$p(\mathbf{B}|\mathbf{D},\mathbf{A}) = \textstyle\prod_{i=1}^{n} p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) \tag{26}$$

$$p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) = \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}}\exp[-(1/2)(\mathbf{b}_i - \mathbf{m}_i(\mathbf{A}))'(d_{ii}\mathbf{M}_i)^{-1}(\mathbf{b}_i - \mathbf{m}_i(\mathbf{A}))]. \tag{27}$$

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 $\phi\mathbf{a}_i$.[10] 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

[10] 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 $\mathbf{A}^{-1}$: $\mathbf{y}_t = \phi\mathbf{y}_{t-1} + \varepsilon_t$.

constant term (the last element of $\mathbf{b}_i$); for details see Appendix A.

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.[11] 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 (28) and (29). We set $V_i = 0.1$, which gives this prior information a weight roughly equivalent to 3 observations; again see Appendix A 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.

5.5 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

[11] 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.

in Figure 1. The data turn out to be quite informative about the values of ψ^y, ψ^π and β^d but cause more modest revisions in our beliefs about other parameters.

Figure 2 highlights the features of the data that lead to these results. Consider simple maximum likelihood estimation with the contemporaneous coefficients parameterized as in (20) with no prior information. Given a particular numerical value for ζ^y and ζ^π, the model would be just-identified and we could calculate maximum likelihood estimates of the other three contemporaneous coefficients α^s, β^d, and γ^d. The figure summarizes the sign of the maximum likelihood estimates $\hat{\alpha}(\zeta^y, \zeta^\pi)$ and $\hat{\beta}(\zeta^y, \zeta^\pi)$ implied by different choices of ζ^y and ζ^π; see Appendix B for the details behind these calculations. The MLE $\hat{\alpha}(\zeta^y, \zeta^\pi)$ would be positive at any point within the shaded regions, and varies from 0 to $+\infty$ as we move from the dashed line to the dotted line within a shaded region. Thus if we had only the sign restrictions on ζ^y, ζ^π, and α^s, the data would rule out combinations for which one of the ζ's is small and the other large but put no restrictions on α^s.

The MLE $\hat{\beta}(\zeta^y, \zeta^\pi)$ is positive in the light-shaded regions of the figure and negative in the dark-shaded regions. It turns out that there are very few values of ζ^y and ζ^π below 0.6 for which both $\alpha^s > 0$ (as required) and $\beta^d > 0$ (as expected by our prior beliefs). If we had imposed a dogmatic prior that required $\beta^d > 0$, the set of maximum likelihood estimates for ζ^y and ζ^π would consist of the light shaded areas in Figure 2. These define a disjoint region with a very odd topography most of whose values are deemed by our prior to be relatively unlikely. This figure highlights a problem with insisting that a parameter like β^d has to be positive. By contrast, our approach nudges the posterior in the direction of favoring the light shaded regions, but also weighs this against the prior plausibility of the values for α^s, ζ^y, and ζ^π that these parameter values would imply. Since the values for these parameters that would result from a positive value of β^d are regarded as relatively unlikely a priori, our posterior inference ends up putting considerable posterior probability on the dark shaded regions in Figure 2 for which $\beta^d < 0$. In other words, contrary to our prior expectation, there is some evidence in the data that higher inflation lowers aggregate demand even with the nominal interest rate held constant.

Impulse-response functions are plotted in Figure 3. The solid lines plot the mean 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 3 represent 95% posterior credibility regions.

The first column of Figure 3 summarizes the effects of a supply shock, which raises output and lowers inflation, with the Fed's choice of interest rate in response to this combination unclear. The second column 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. 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 0.57% and inflation falling 0.24%, as a result of which in equilibrium the fed funds rate is only 31 basis points higher in the immediate response to the shock. The output effect declines in the quarters following the shock, with the posterior mean actually switching signs after 6 quarters.

Figure 4 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 gap (in deviations from the sample mean). The solid line in the top panel is the contribution of supply shocks over the 10 years prior to the indicated date,[12] while the second and third panels give the contributions of demand and monetary policy shocks, respectively. The shaded regions denote 95% posterior credibility regions; to our knowledge, this is the first time such error bands have ever been reported in the very large literature using VARs 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 5 and 6 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

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

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 2 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 91% of the variance of interest rates and supply shocks account for about 2/3 of the variance of inflation. Demand shocks account for half of the variability of output and supply shocks another third. Monetary policy shocks are significantly less important for determining the paths of any of the three variables.

6 Conclusion.

Structural inference is only possible if we have prior information about the underlying economic model and mechanisms. It is possible to represent this information in a less dogmatic way than is done with conventional identifying assumptions. If we do so, the model is 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 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

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2016). "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," working paper, Emory University.

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 (2016). "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," working paper, UCSD.

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.

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 (2016a), "Delta-Method Inference for a Class of Set-Identified SVARs," working paper, New York University.

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

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.

Moon, Hyungsik Roger, Frank Schorfheide, Eleonora Granziera, and Mihye Lee (2011). "Inference for VARs Identified with Sign Restrictions," NBER Working Paper 17140.

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.

Appendix A. 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

$$\underset{[(T+13+1)\times 1]}{\tilde{\mathbf{Y}}_i} = \begin{bmatrix} \mathbf{a}_i'\mathbf{y}_1 & \cdots & \mathbf{a}_i'\mathbf{y}_T & \mathbf{m}_i'\mathbf{P}_i & r_i/\sqrt{V_i} \end{bmatrix}' \tag{28}$$

$$\underset{[(T+13+1)\times 13]}{\tilde{\mathbf{X}}_i} = \begin{bmatrix} \mathbf{x}_0 & \cdots & \mathbf{x}_{T-1} & \mathbf{P}_i & \mathbf{e}_i/\sqrt{V_i} \end{bmatrix}' \tag{29}$$

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 0.2, 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 (28) and (29). 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 (29) to be column 3 of $\mathbf{I}_{13}$ and r_i in equation (28) 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.

Appendix B. Derivation of Figure 2.

The maximum likelihood estimate of the reduced-form innovation variance matrix is

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} 0.1835 & -0.0137 & 0.0524 \\ -0.0137 & 0.1228 & 0.0280 \\ 0.0524 & 0.0280 & 0.0972 \end{bmatrix}.$$

Notice that for any given values of ζ^y and ζ^π, the maximum likelihood estimate of α^s could be found from an IV regression of $\hat{\varepsilon}_t^y$ on $\hat{\varepsilon}_t^\pi$ using $(\hat{\varepsilon}_t^r - \zeta^y \hat{\varepsilon}_t^y - \zeta^\pi \hat{\varepsilon}_t^\pi)$ as instrument,

$$\hat{\alpha}(\zeta^y, \zeta^\pi) = \frac{\sum_{t=1}^T (\hat{\varepsilon}_t^r - \zeta^y \hat{\varepsilon}_t^y - \zeta^\pi \hat{\varepsilon}_t^\pi)\hat{\varepsilon}_t^y}{\sum_{t=1}^T (\hat{\varepsilon}_t^r - \zeta^y \hat{\varepsilon}_t^y - \zeta^\pi \hat{\varepsilon}_t^\pi)\hat{\varepsilon}_t^\pi} = \frac{\hat{\omega}_{13} - \zeta^y \hat{\omega}_{11} - \zeta^\pi \hat{\omega}_{12}}{\hat{\omega}_{23} - \zeta^y \hat{\omega}_{21} - \zeta^\pi \hat{\omega}_{22}}, \tag{30}$$

for $\hat{\omega}_{ij}$ the (i, j) element of $\hat{\boldsymbol{\Omega}}$.

Figure 2 plots possible values for ζ^y on the horizontal axis and ζ^π on the vertical axis. The steeply sloping dashed line identifies combinations of ζ^y and ζ^π for which the numerator of (30) would be zero, that is, the dashes plot the line $\hat{\omega}_{13} - \zeta^y \hat{\omega}_{11} - \zeta^\pi \hat{\omega}_{12} = 0$. This line could alternatively be described as combinations of ζ^y and ζ^π for which the structural monetary policy shock ($u_t^m = \varepsilon_t^r - \zeta^y \varepsilon_t^y - \zeta^\pi \varepsilon_t^\pi$) would be uncorrelated with the reduced-form residual for output (ε_t^y) and therefore for which the MLE of α^s (the value that would make $\varepsilon_t^y - \alpha^s \varepsilon_t^\pi$ uncorrelated with u_t^m) is in fact zero. The flatter dotted line plots combinations of ζ^y and ζ^π for which the denominator of (30) is zero, namely $\hat{\omega}_{23} - \zeta^y \hat{\omega}_{21} - \zeta^\pi \hat{\omega}_{22} = 0$. These are values that would make the structural monetary shock uncorrelated with the reduced-form residual for inflation and would imply an infinite value for the MLE of α^s.

In order for α^s to be positive, the numerator and denominator of (30) would have to be of the same sign. Since $\hat{\omega}_{12} < 0$, the numerator is positive for any pair (ζ^y, ζ^π) above the dashed line and negative for any point below the dashed line. The denominator of (30) is positive for any point below the dotted line. Thus in order to satisfy $\alpha^s > 0$, the values of ζ^y and ζ^π would have to be in a shaded region of Figure 2, either both above the dashed line and below the dotted (the lower left quadrant of Figure 2) or below the dashed and above the dotted (the upper right quadrant). As one moves from the dashed line to the dotted line within a shaded region, the MLE of α^s would vary from 0 to $+\infty$.

Given the maximum likelihood estimate $\hat{\alpha}(\zeta^y, \zeta^\pi)$ associated with any specified (ζ^y, ζ^π), we can find the MLE for β^d and γ^d associated with that (ζ^y, ζ^π) by an IV regression of $\hat{\varepsilon}_t^y$ on $\hat{\varepsilon}_t^\pi$ and $\hat{\varepsilon}_t^r$ using $(\hat{\varepsilon}_t^y - \hat{\alpha}(\zeta^y, \zeta^\pi))$ and $(\hat{\varepsilon}_t^r - \zeta^y \hat{\varepsilon}_t^y - \zeta^\pi \hat{\varepsilon}_t^\pi)$ as instruments:

$$\begin{bmatrix} \hat{\beta}(\zeta^y, \zeta^\pi) \\ \hat{\gamma}(\zeta^y, \zeta^\pi) \end{bmatrix} = \begin{bmatrix} [\hat{\omega}_{12} - \hat{\alpha}(\zeta^y, \zeta^\pi)\hat{\omega}_{22}] & [\hat{\omega}_{13} - \hat{\alpha}(\zeta^y, \zeta^\pi)\hat{\omega}_{23}] \\ [\hat{\omega}_{32} - \zeta^y \hat{\omega}_{12} - \zeta^\pi \hat{\omega}_{22}] & [\hat{\omega}_{33} - \zeta^y \hat{\omega}_{13} - \zeta^\pi \hat{\omega}_{23}] \end{bmatrix}^{-1} \times \begin{bmatrix} [\hat{\omega}_{11} - \hat{\alpha}(\zeta^y, \zeta^\pi)\hat{\omega}_{12}] \\ [\hat{\omega}_{31} - \zeta^y \hat{\omega}_{11} - \zeta^\pi \hat{\omega}_{21}] \end{bmatrix}. \tag{31}$$

One can use these expressions to calculate combinations of ζ^y and ζ^π that would imply a negative value for β^d. This corresponds to the region below the beaded curve and above the solid curve in Figure 2, and appears as dark shaded regions provided that the conditions for $\alpha^s > 0$ are satisfied. There are a number of points at which the matrix in (31) becomes singular and no values for β^d and γ^d could achieve the same fit as the reduced-form VAR. As one crosses these points, signs of key relations flip and the markers denoting a given curve changes status from solid (the lower bound on the region) to beaded (the upper bound).

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	1.5	0.4	none
γ^d	Effect of r on demand	-2	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 $\alpha = 13$ and $\beta = 5$		
ρ	Interest rate smoothing	0.75	0.1	$0 \leq \rho \leq 1$

Table 2. Decomposition of variance of 4-quarter-ahead forecast errors.

	Supply	Demand	Monetary policy
Output gap	**0.42** [33%] (0.06, 1.02)	**0.63** [50%] (0.28, 1.17)	**0.22** [17%] (0.05, 0.53)
Inflation	**0.48** [67%] (0.22, 0.88)	**0.15** [21%] (0.02, 0.37)	**0.09** [13%] (0.01, 0.23)
Fed funds rate	**0.06** [3%] (0.00, 0.27)	**1.56** [91%] (0.90, 2.53)	**0.10** [6%] (0.01, 0.38)

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

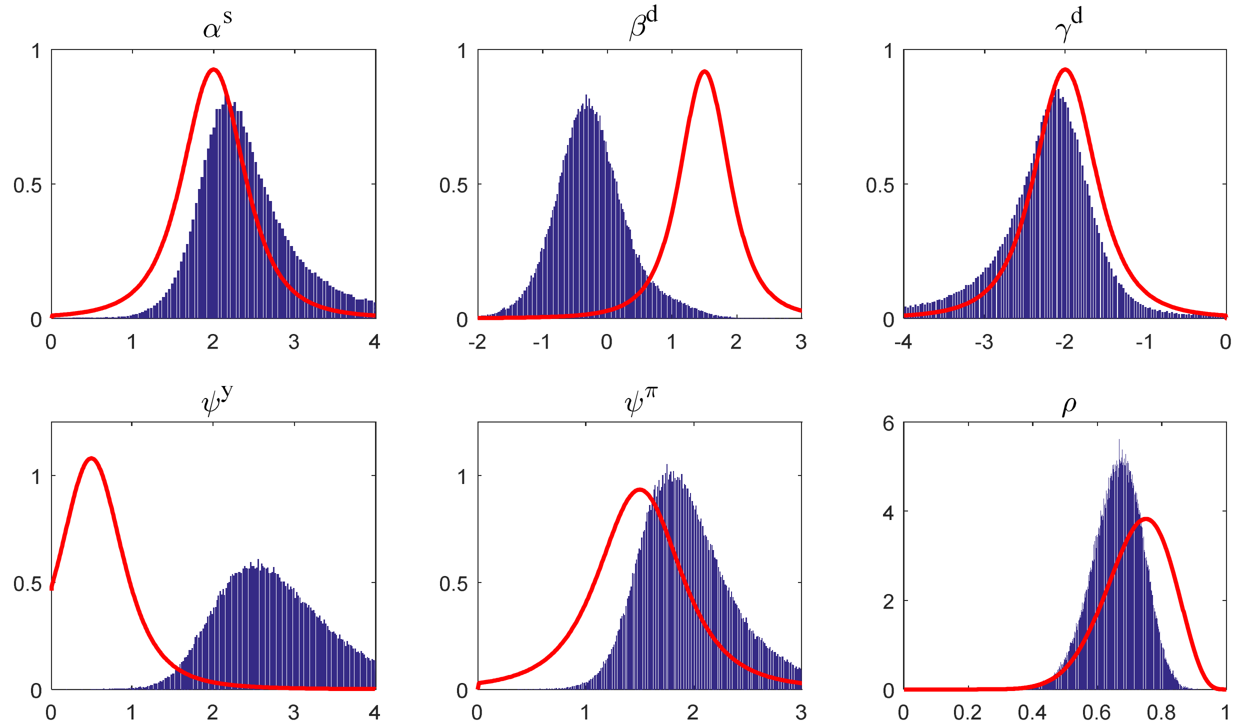

Figure 1. Prior distributions (red lines) and posterior distributions (blue histogram) for contemporaneous coefficients.

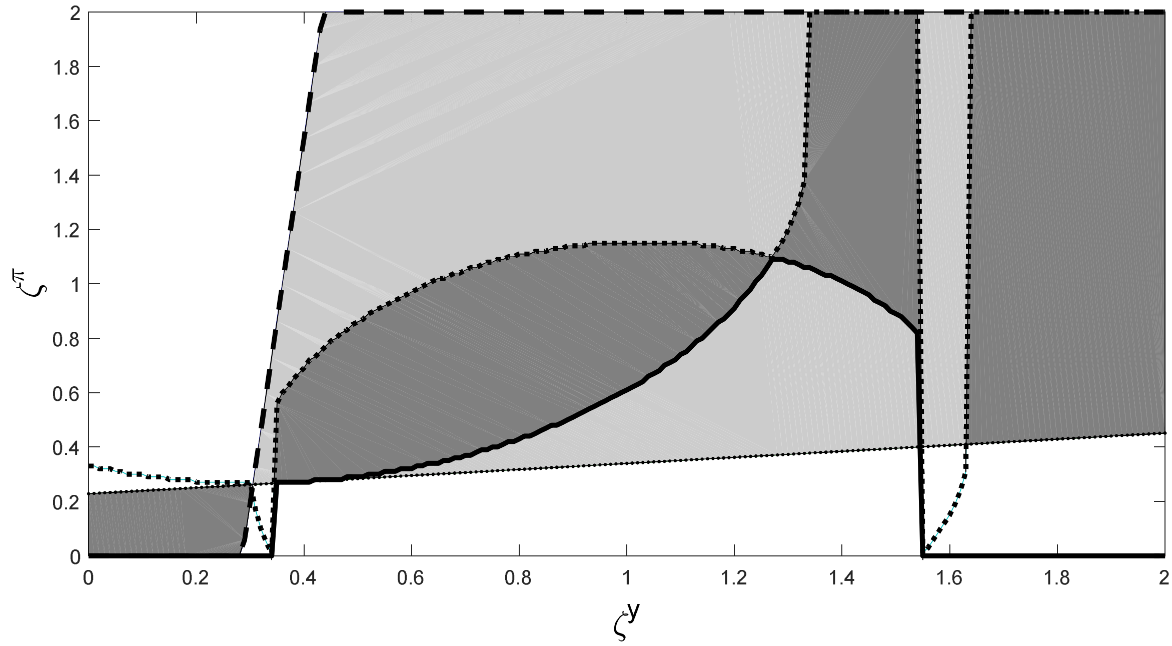

Figure 2. Implications of different values for ζ^y and ζ^π. Dashed line: values for which MLE of α^s is zero. Dotted line: values for which MLE of α^s is infinity. Dark shaded region: $\alpha^s > 0$ and $\beta^d < 0$. Light shaded region: $\alpha^s > 0$ and $\beta^d > 0$.

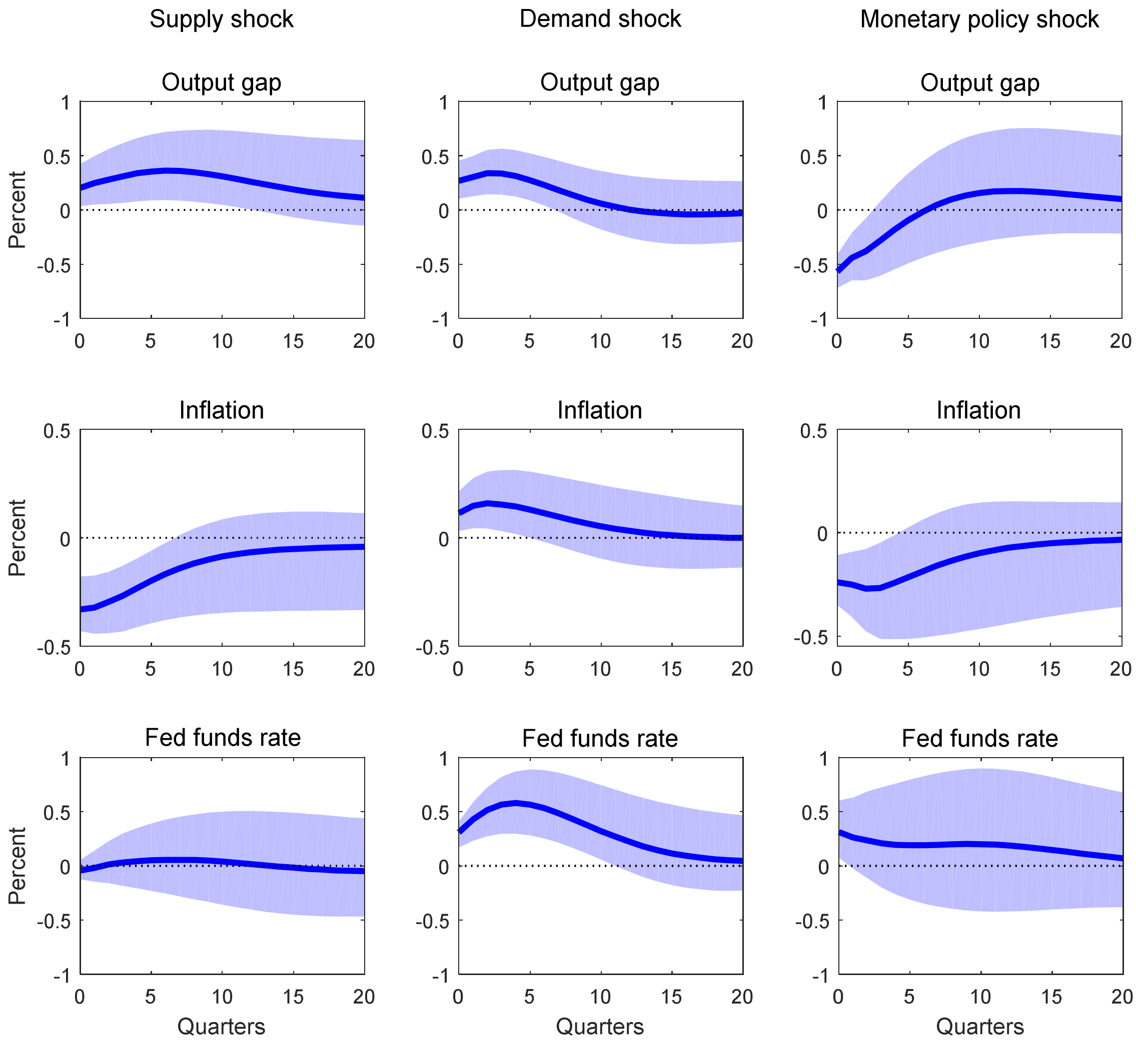

Figure 3. Impulse-response functions for 3-variable VAR with informative priors for lagged coefficients. Solid lines: posterior mean. Shaded regions: 95% posterior credibility set.

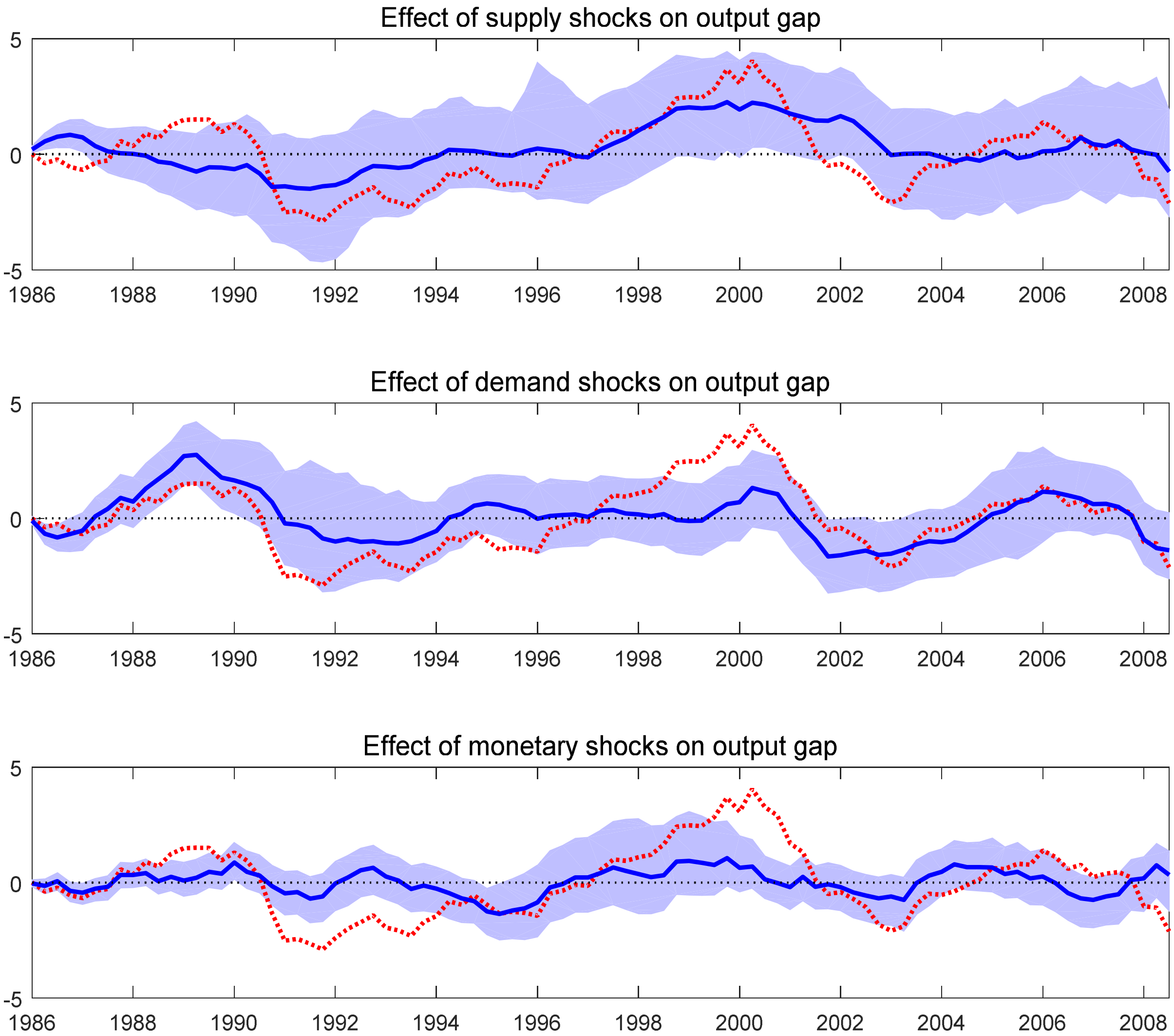

Figure 4. Portion of historical variation in output gap attributed to each of the structural shocks. Dotted red: actual value for the deviation of output gap from its mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 95% posterior credibility sets.

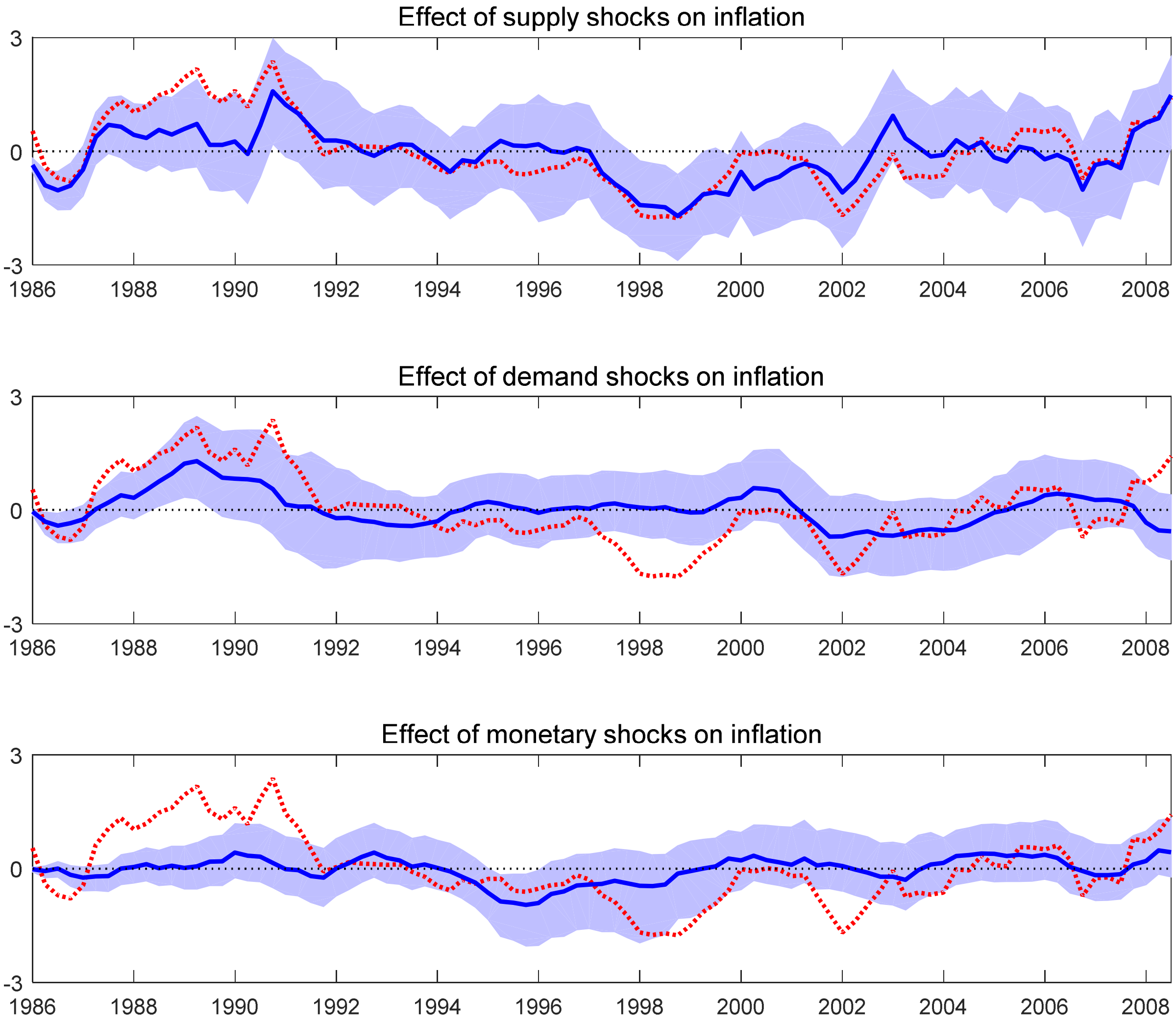

Figure 5. Portion of historical variation in inflation attributed to each of the structural shocks. Dotted red: actual value for the deviation of inflation from its mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 95% posterior credibility sets.

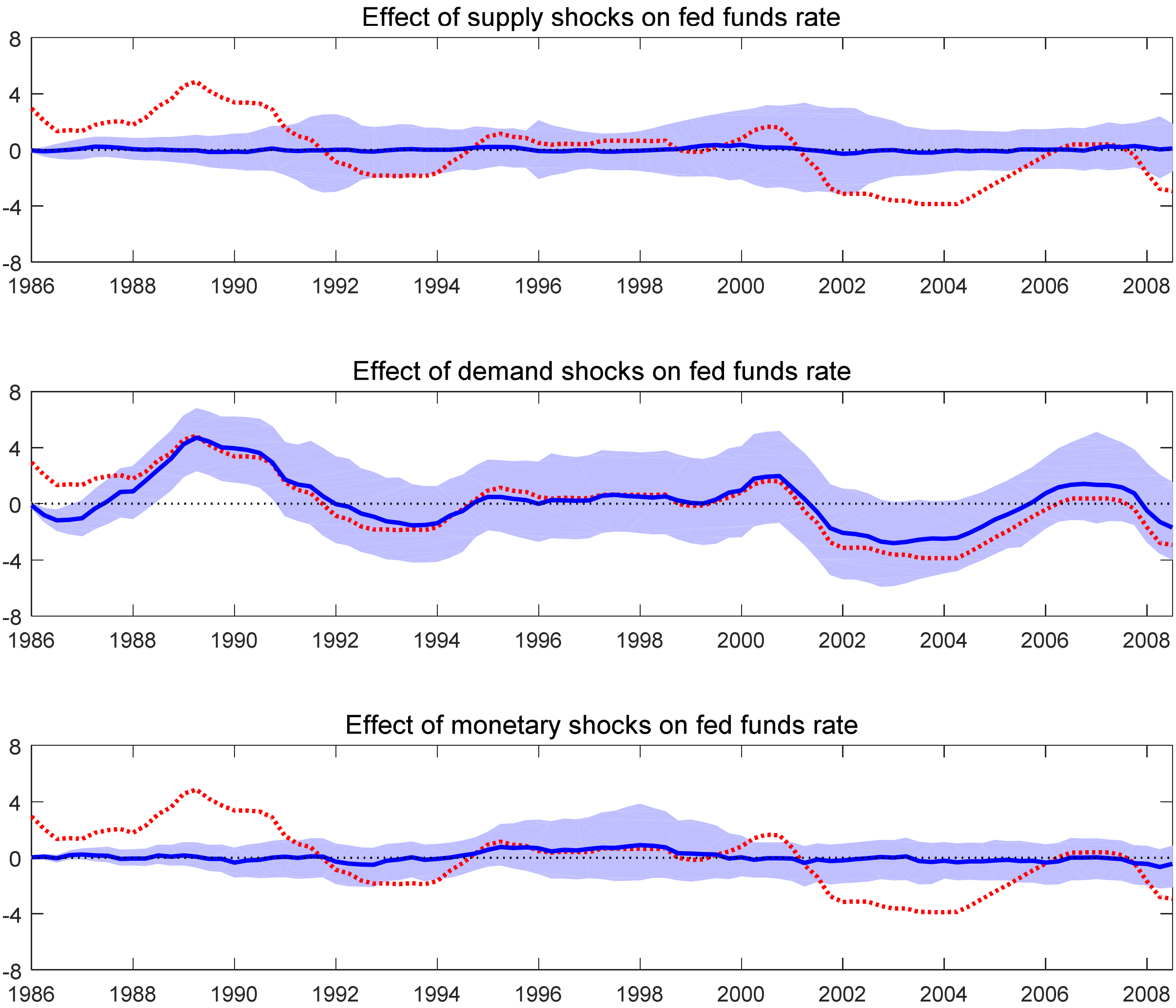

Figure 6. Portion of historical variation in fed funds rate attributed to each of the structural shocks. Dotted red: actual value for the deviation of fed funds rate from its mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 95% posterior credibility sets.