Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information*

Christiane Baumeister

Bank of Canada

cbaumeister@bankofcanada.ca

James D. Hamilton

University of California at San Diego

jhamilton@ucsd.edu

November 18, 2013

Revised: March 7, 2014

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada. We thank Fabio Canova, Gary Chamberlain, Drew Creal, Lutz Kilian and Adrian Pagan for helpful comments.

ABSTRACT

Many empirical studies have used numerical Bayesian methods for structural inference in vector autoregressions that are identified solely on the basis of sign restrictions. Because sign restrictions only provide set-identification of structural parameters, over certain regions of the parameter space the posterior inference could only be a restatement of prior beliefs. In this paper we characterize these regions, explicate the beliefs about parameters that are implicit in conventional priors, provide an analytical characterization of the full posterior distribution for arbitrary priors, and analyze the asymptotic properties of this posterior distribution. We show that in a bivariate supply and demand example, if the population correlation between the VAR residuals is negative, then even if one has available an infinite sample of data, any inference about the supply elasticity is coming solely from the prior distribution. More generally, the asymptotic posterior distribution of contemporaneous coefficients in an n-variable VAR is confined to the set of values that orthogonalize the population variance-covariance matrix of OLS residuals, with the height of the posterior proportional to the height of the prior at any point within that set. We suggest that researchers should defend their prior beliefs explicitly and report the difference between prior and posterior distributions for key magnitudes of interest. We illustrate these methods with a simple macroeconomic model.

1 Introduction.

In pioneering papers, Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) proposed that structural inference using vector autoregressions might be based solely on prior beliefs about the signs of the impacts of certain shocks. This approach has since been adopted in hundreds of follow-up studies, and today is one of the most popular tools used by researchers who seek to draw structural conclusions using VARs.

However, structural inference using sign restrictions is not without its shortcomings. An assumption about signs is not enough by itself to identify structural parameters. What the procedure actually delivers is a set of possible inferences, each of which is equally consistent with both the observed data and the underlying restrictions. This feature raises a number of challenges for users of the method.

From a frequentist perspective, the task of describing the set of values for magnitudes of interest that cannot be rejected on the basis of classical hypothesis tests is both computationally demanding and conceptually awkward (see for example Moon, Schorfheide, and Granziera, 2013). For this reason, the vast majority of applications use Bayesian methods. However, the Bayesian approach raises another set of troubling issues. Most important among these is the observation by Poirier (1998) and Moon and Schorfheide (2012) that, since the data are uninformative about certain regions of the parameter space, for some questions the Bayesian posterior inference will be determined purely from prior beliefs even if the sample size is infinite. This key feature is overlooked in most of the published studies using sign-restricted VARs.

We advocate that instead researchers should base inference on informative priors that reflect not just sign restrictions but also the relative plausibility of different parameter values within the allowable set. We develop algorithms for Bayesian inference in structural vector autoregressions with informative priors, generalizing the results in Sims and Zha (1998) to a setting where some parameters may only be set-identified. Our formulation also takes advantage of a natural conjugate distribution for structural variances to simplify the computational demands of Sims and Zha's formulas. A key object that we suggest that researchers should report is the difference between the prior and posterior distributions. Looking at these can help get beyond the "black box" character of many sign-restriction studies and allow the researcher to elucidate the exact features of the observed data that have mattered for informing the posterior inference.

We demonstrate that as the sample size goes to infinity, the analyst could know with certainty that structural parameters fall within a set $S(\mathbf{\Omega})$ that orthogonalizes the true variance-covariance matrix, but within this set, the height of the infinite-sample posterior distribution is simply a constant times the height of the prior distribution at that point. In the case of a bivariate model in which sign restrictions are the sole identifying assumption, if the reduced-form residuals have negative correlation, then $S(\mathbf{\Omega})$ allows any value for the elasticity of supply but restricts the elasticity of demand to fall within a particular interval. With positively correlated errors, the elasticity of demand could be any negative number while the elasticity of supply is restricted to fall in a particular interval. We also explore the implications of the popular uniform Haar prior. We show that although this prior is uninformative about the angles of rotation in certain matrices, for magnitudes of interest such as elasticities, the Haar prior favors some values over others. The empirical conclusions that are typically reported reflect these implicit prior beliefs in ways that have not been recognized by applied researchers.

We claim a number of other benefits to our approach as well. First, given the weak nature of information captured by sign restrictions alone, substantial gains may be available from using additional information. For example, we may be confident not just that a supply elasticity is positive but further may regard extremely large values as implausible. Kilian and Murphy (2012) argued persuasively that incorporating such additional prior information can substantially improve the inference. Their approach calls for insisting that supply elasticities above a certain maximum can be ruled out a priori, but simultaneously using no information about the plausibility of positive values just below that maximum. In practice, researchers like Kilian and Murphy (2012) and Juvenal and Petrella (forthcoming) use a set of different possible upper bounds as a form of robustness checking, acknowledging as a practical matter that we're not sure exactly where to impose the cutoffs. A more natural approach would allow the prior plausibility to decline gradually as one considers higher values for the supply elasticity, approaching zero continuously rather than exhibiting a discrete drop at some cutoff. In this paper, we provide algorithms for Bayesian inference using arbitrary priors for supply elasticities and other parameters, including prior densities with abrupt, discrete cutoffs as well as priors that are continuous functions of the unknown parameters. In a related paper, Caldara and Kamps (2012) nicely illustrated how one can use prior information about the progressivity of taxes to draw conclusions about the effects of changes in tax laws on output. The tools developed here allow incorporating the insights of Kilian and Murphy (2012) and Caldara and Kamps (2012) for use in any structural vector autoregression for which the researcher may have more confidence in some prior restrictions than others.

Second, our approach provides a constructive resolution of a key ambiguity in the literature using sign-restricted VARs. Applied researchers invariably want to report their results in terms of point estimates. But that is problematic if the method only identifies a set of possible answers. Fry and Pagan (2011) observed that it makes no sense from a frequentist perspective to calculate the median across a set of equivalent models and that the point estimates that are typically reported are in fact not consistent with the model's theoretical restriction that structural shocks should be uncorrelated. Inoue and Kilian (2013) proposed one solution to this problem based on Bayesian analysis using the uniform Haar prior, calculating the joint posterior distribution of the set of impulse-response coefficients and reporting the path with the highest posterior mode. By contrast, our suggestion is that the researcher should represent subjective beliefs in the form of an informative prior distribution. The output of the statistical analysis is then a posterior distribution that summarizes what has been learned from seeing the data. If we are interested in a particular magnitude, such as the response after s periods of variable i to the jth structural shock, and specify a loss function that summarizes the cost of getting the answer to that question wrong, then there is an unambiguous optimal estimate to report. For example, with a quadratic loss function, the optimal point estimate is the posterior mean of $\partial y_{i,t+s}/\partial u_{jt}$, though we demonstrate below analytically that the expected loss may be infinite and the posterior mean may fail to exist when the uniform Haar prior is used. If the loss function is the absolute value of the deviation between the estimate and the true value, then we should report the posterior median. Our procedure answers the Fry-Pagan objections in that the researcher reports the full joint probability distribution over the set of all models, and any point estimates reported represent the optimal estimate of the magnitude of interest. The calculation that delivers this estimate is obtained not by an arbitrary ranking of equivalent models but instead by integration over a well-defined probability distribution.

Indeed, with a well posed prior it is possible to calculate the unique optimal posterior inference about any magnitude of interest. Such results are helpful not just for Bayesian inference but for frequentist analysis as well, because any frequentist procedure that cannot be derived as the optimal Bayesian decision for some prior and loss function should be possible to improve upon.[1]

Finally, we note that our method allows the researcher to put as much or as little weight on the prior as desired and to report how the results would change as a result. Our suggestion is that researchers might typically want to report how the posterior inference changes as the prior becomes less informative, as a road map for assessing the robustness of inference and understanding the mapping between likelihood, prior, and posterior as advocated by Leamer

[1] See for example Ferguson (1967, Chapter 2), Müller and Norets (2012), and Elliott, Müller, and Watson (2012).

(1978).[2]

We illustrate the promise of this method with a simple three-variable macroeconomic model. We find that 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. The data are also modestly informative about Taylor Rule parameters governing the response of the Federal Reserve to the output gap and inflation, suggesting a smaller Fed response to inflation than anticipated a priori. Overall, after seeing the data, a researcher would be more confident that a monetary contraction lowers output and inflation, although there is no strong evidence of the output effect lasting more than a few quarters.

The plan of the paper is as follows. Section 2 describes a possibly set-identified n-variable VAR and derives the Bayesian posterior distribution for an arbitrary prior distribution on contemporaneous coefficients assuming that priors for other parameters are chosen from the natural conjugate classes. We also analyze the asymptotic properties of Bayesian inference in this general setting. Section 3 discusses the use of sign restrictions for impacts at longer horizons, and suggests that the correct way to approach these from a Bayesian perspective is in the form of beliefs about the interaction between contemporaneous and lagged structural coefficients. Section 4 illustrates these results in the case of a simple 2-variable example based on supply and demand and relates them to the traditional approach to sign restrictions. Section 5 applies our recommended approach to a 3-variable macroeconomic model. Section

[2] Giacomini and Kitagawa (2013) proposed forming priors directly on the set of orthogonal matrices that could transform residuals orthogonalized by the Cholesky factorization into an alternative orthogonalized structure, and investigate the sensitivity of the resulting inference to the priors. By contrast, our approach is to formulate priors directly in terms of beliefs about the economic structure.

6 briefly concludes.

2 Bayesian inference for partially identified structural vector autoregressions.

We investigate dynamic 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, $\mathbf{A}$ 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 $\mathbf{u}_t$ an $(n \times 1)$ vector of structural disturbances assumed to be i.i.d. $N(\mathbf{0}, \mathbf{D})$ and mutually uncorrelated ($\mathbf{D}$ diagonal). The reduced-form VAR associated with the structural model (1) is given by

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

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

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

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t) = \mathbf{\Omega} = \mathbf{A}^{-1}\mathbf{D}(\mathbf{A}^{-1})'. \tag{5}$$

In this section we suppose that the investigator begins with prior beliefs about the values of the structural parameters represented by a density $p(\mathbf{A}, \mathbf{D}, \mathbf{B})$, and show how observation of the data $\mathbf{Y}_T = (\mathbf{x}'_0, \mathbf{y}'_1, \mathbf{y}'_2, ..., \mathbf{y}'_T)'$ would lead the investigator to revise those beliefs.[3]

[3] Our derivations draw on insights from Sims and Zha (1998). The main difference is that we parameterize the contemporaneous relations in terms of two matrices $\mathbf{A}$ and $\mathbf{D}$, whereas they use a single matrix. This allows us to derive simpler expressions and closed-form results along with asymptotic properties of Bayesian inference.

We represent prior information about the contemporaneous structural coefficients in the form of an arbitrary prior distribution $p(\mathbf{A})$. This prior could incorporate any combination of exclusion restrictions, sign restrictions, and informative prior beliefs about elements of $\mathbf{A}$. For example, our procedure could be used to calculate the posterior distribution even if no sign or exclusion restrictions were imposed. We also allow for interaction between the prior beliefs about different parameters by specifying conditional prior distributions $p(\mathbf{D}|\mathbf{A})$ and $p(\mathbf{B}|\mathbf{A}, \mathbf{D})$ that potentially depend on $\mathbf{A}$. We assume that there are no restrictions on the lag coefficients in $\mathbf{B}$ other than the prior beliefs represented by the distribution $p(\mathbf{B}|\mathbf{A}, \mathbf{D})$.

To represent possible prior information about $\mathbf{D}$ and $\mathbf{B}$, we employ natural conjugate priors to facilitate analytical characterization of results as well as to allow for simple empirical implementation. We use $\Gamma(\kappa_i, \tau_i)$ priors for the reciprocals of diagonal elements of $\mathbf{D}$, taken to be independent across equations,[4]

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

$$p(d_{ii}^{-1}|\mathbf{A}) = \begin{cases} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}(d_{ii}^{-1})^{\kappa_i - 1}\exp(-\tau_i d_{ii}^{-1}) & \text{for } d_{ii}^{-1} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where d_{ii} denotes the row i, column i element of $\mathbf{D}$. Note that κ_i/τ_i denotes the prior mean for d_{ii}^{-1} and κ_i/τ_i^2 its variance.

Normal priors are used for the lagged structural coefficients $\mathbf{B}$, with results particularly

[4] We will follow the notational convention of using $p(.)$ to denote an arbitrary density, with the density being referred to implicit by the argument. Thus $p(\mathbf{A})$ is shorthand notation for $p_{\mathbf{A}}(\mathbf{A})$ and represents a different function from $p(\mathbf{D})$, which in more careful notation would be denoted $p_{\mathbf{D}}(\mathbf{D})$.

simple if coefficients are taken to be independent $N(\mathbf{m}_i, d_{ii}\mathbf{M}_i)$ across equations:

$$p(\mathbf{B}|\mathbf{D},\mathbf{A}) = \textstyle\prod_{i=1}^{n} p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) \tag{7}$$

$$p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) = \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}} \exp[-(1/2)(\mathbf{b}_i - \mathbf{m}_i)'(d_{ii}\mathbf{M}_i)^{-1}(\mathbf{b}_i - \mathbf{m}_i)]. \tag{8}$$

Here $\mathbf{b}_i'$ denotes the ith row of $\mathbf{B}$ (the lagged coefficients for the ith structural equation). Thus $\mathbf{m}_i$ denotes the prior mean for the lagged coefficients in the ith equation and $d_{ii}\mathbf{M}_i$ denotes the variance associated with this prior. In the specific examples employed below we pose the prior in terms of an expected value $\boldsymbol{\eta}$ for the reduced-form coefficient matrix $\boldsymbol{\Phi}$ which implies that $\mathbf{m}_i' = \mathbf{a}_i'\boldsymbol{\eta}$ where $\mathbf{a}_i'$ denotes the ith row of $\mathbf{A}$. The overall prior is thus

$$p(\mathbf{A},\mathbf{D},\mathbf{B}) = p(\mathbf{A}) \textstyle\prod_{i=1}^{n} [p(d_{ii}|\mathbf{A})p(\mathbf{b}_i|\mathbf{D},\mathbf{A})]. \tag{9}$$

With Gaussian residuals, the likelihood function (conditioning on the pre-sample values of $\mathbf{y}_0, \mathbf{y}_{-1}, ..., \mathbf{y}_{-m+1}$) is given by

$$\begin{aligned} p(\mathbf{Y}_T|\mathbf{A},\mathbf{D},\mathbf{B}) &= (2\pi)^{-Tn/2}|\det(\mathbf{A})|^T|\mathbf{D}|^{-T/2} \times \\ & \quad \exp\left[-(1/2)\textstyle\sum_{t=1}^{T}(\mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1})'\mathbf{D}^{-1}(\mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1})\right] \end{aligned} \tag{10}$$

where $|\det(\mathbf{A})|$ denotes the absolute value of the determinant of $\mathbf{A}$.

In Appendix A we derive the following characterization of the posterior distribution and detail in Appendix B an algorithm that can be used to generate draws from this distribution.

Proposition 1. *Let* $\mathbf{a}_i'$ *denote the ith row of* $\mathbf{A}$, $\phi(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ *denote the multivariate Normal density with mean* $\boldsymbol{\mu}$ *and variance* $\boldsymbol{\Sigma}$ *evaluated at* $\mathbf{x}$ *and* $\gamma(x;\kappa,\tau)$ *denote a gamma density with parameters* κ *and* τ *evaluated at* x. *If the likelihood is (10) and priors are* *given by (6)-(9) with* $\mathbf{m}_i' = \mathbf{a}_i'\boldsymbol{\eta}$, *then the posterior distribution can be written as*

$$p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T) = p(\mathbf{A}|\mathbf{Y}_T)p(\mathbf{D}|\mathbf{A}, \mathbf{Y}_T)p(\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T)$$

with

$$p(\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T) = \textstyle\prod_{i=1}^n \phi(\mathbf{b}_i; \mathbf{m}_i^*, d_{ii}\mathbf{M}_i^*)$$

$$\mathbf{m}_i^* = \left(\textstyle\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{x}_{t-1}' + \mathbf{M}_i^{-1}\right)^{-1} \left(\textstyle\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{y}_t'\mathbf{a}_i + \mathbf{M}_i^{-1}\mathbf{m}_i\right) \tag{11}$$

$$\mathbf{M}_i^* = \left(\textstyle\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{x}_{t-1}' + \mathbf{M}_i^{-1}\right)^{-1} \tag{12}$$

$$p(\mathbf{D}|\mathbf{A}, \mathbf{Y}_T) = \textstyle\prod_{i=1}^n \gamma(d_{ii}^{-1}; \kappa_i^*, \tau_i^*)$$

$$\kappa_i^* = \kappa_i + (T/2) \tag{13}$$

$$\tau_i^* = \tau_i + (T/2)\mathbf{a}_i'\hat{\boldsymbol{\Omega}}_{iT}^*\mathbf{a}_i \tag{14}$$

$$\begin{aligned} \hat{\boldsymbol{\Omega}}_{iT}^* &= T^{-1}\Big\{\textstyle\sum_{t=1}^T \mathbf{y}_t\mathbf{y}_t' + \boldsymbol{\eta}\mathbf{M}_i^{-1}\boldsymbol{\eta}' \\ &\quad - \left(\textstyle\sum_{t=1}^T \mathbf{y}_t\mathbf{x}_{t-1}' + \boldsymbol{\eta}\mathbf{M}_i^{-1}\right)\mathbf{M}_i^*\left(\textstyle\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{y}_t' + \mathbf{M}_i^{-1}\boldsymbol{\eta}'\right)\Big\} \end{aligned} \tag{15}$$

$$\hat{\boldsymbol{\Omega}}_T^* = n^{-1}\textstyle\sum_{i=1}^n \hat{\boldsymbol{\Omega}}_{iT}^* \tag{16}$$

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\boldsymbol{\Omega}}_T^*\mathbf{A}')]^{T/2}}{\prod_{i=1}^n [(2\tau_i/T) + \mathbf{a}_i'\hat{\boldsymbol{\Omega}}_{iT}^*\mathbf{a}_i]^{\kappa_i+T/2}} \tag{17}$$

where k_T *denotes the constant for which (17) integrates to unity.*

Consider first the posterior distribution for $\mathbf{b}_i$, the lagged coefficients in the ith structural equation, conditional on $\mathbf{A}$ and $\mathbf{D}$. In the special case of a noninformative prior for these coefficients ($\mathbf{M}_i^{-1} = \mathbf{0}$), this takes the form of a Normal distribution centered at $\mathbf{m}_i^* = \left(\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}\left(\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{y}_t'\mathbf{a}_i\right)$, or the coefficient from an OLS regression of $\mathbf{a}_i'\mathbf{y}_t$ on

$\mathbf{x}_{t-1}$, with variance given by $d_{ii}\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}$, again the OLS formula. Although the Bayesian would describe $\mathbf{m}_i^*$ and $d_{ii}\mathbf{M}_i^*$ as moments of the posterior distribution, they are simple functions of the data, and it is also straightforward to use a frequentist perspective to summarize the properties of the Bayesian posterior inference. In particular, as long as $\mathbf{M}_i^{-1}$ is finite and the true process for $\mathbf{y}_t$ is covariance-stationary and ergodic for second moments, we have that as the sample size T gets large,

$$\begin{aligned}\mathbf{m}_i^* &= \left(T^{-1}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}' + T^{-1}\mathbf{M}_i^{-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t'\mathbf{a}_i + T^{-1}\mathbf{M}_i^{-1}\mathbf{m}_i\right)\\ &\xrightarrow{p} \left[E(\mathbf{x}_{t-1}\mathbf{x}_{t-1}')^{-1}\right]E(\mathbf{x}_{t-1}\mathbf{y}_t')\mathbf{a}_i\end{aligned}$$

and $\mathbf{M}_i^* \xrightarrow{p} \mathbf{0}$. In other words, as long as $\mathbf{M}_i^{-1}$ is finite, the values of the prior parameters $\mathbf{m}_i$ and $\mathbf{M}_i$ are asymptotically irrelevant, and the Bayesian posterior distribution for $\mathbf{b}_i$ collapses to a Dirac delta function around the same plim that characterizes the OLS regression of $\mathbf{a}_i'\mathbf{y}_t$ on $\mathbf{x}_{t-1}$. Conditional on $\mathbf{a}_i$, the data are perfectly informative asymptotically about $\mathbf{b}_i$, reproducing the familiar result that, for these features of the parameter space, the Bayesian inference is the same asymptotically as frequentist inference and correctly uncovers the true value.

Similarly for d_{ii}, the variance of the ith structural equation, in the special case of a noninformative prior for $\mathbf{B}$ (that is, when $\mathbf{M}_i^{-1} = \mathbf{0}, i = 1, ..., n$) we have that

$$\begin{aligned}\hat{\mathbf{\Omega}}_{iT}^* &= T^{-1}\left\{\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' - \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}'\right)\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t'\right)\right\}\\ &= \hat{\mathbf{\Omega}}_T \end{aligned}\qquad(18)$$

where $\hat{\mathbf{\Omega}}_T = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$ is the sample variance matrix of $\hat{\boldsymbol{\varepsilon}}_t$, the vector of residuals from

OLS regression of $\mathbf{y}_t$ on $\mathbf{x}_{t-1}$. If the priors for d_{ii} are also noninformative $(\kappa_i = \tau_i = 0)$, then the posterior expectation of d_{ii}^{-1} is given by $\kappa_i^*/\tau_i^* = 1/(\mathbf{a}_i'\hat{\mathbf{\Omega}}_T\mathbf{a}_i)$, the reciprocal of the average squared residual from the OLS regression of $\mathbf{a}_i'\mathbf{y}_t$ on $\mathbf{x}_{t-1}$, with variance $\kappa_i^*/(\tau_i^*)^2 = 2/[T(\mathbf{a}_i'\hat{\mathbf{\Omega}}_T\mathbf{a}_i)^2]$ again shrinking to zero as T gets large. In the case of general but nondogmatic priors (κ_i, τ_i and $\mathbf{M}_i^{-1}$ all finite), as $T \to \infty$, the value of $\hat{\mathbf{\Omega}}_{iT}^*$ still converges to the OLS estimate $\hat{\mathbf{\Omega}}_T$, and the Bayesian posterior distribution for d_{ii}^{-1} conditional on $\mathbf{a}_i$ will collapse to a point mass at $1/(\mathbf{a}_i'\mathbf{\Omega}_0\mathbf{a}_i)$ for $\mathbf{\Omega}_0 = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$ the true variance matrix. Hence again the priors are asymptotically irrelevant for inference about $\mathbf{D}$ conditional on $\mathbf{A}$.

By contrast, prior beliefs about $\mathbf{A}$ will not vanish asymptotically unless the elements of $\mathbf{A}$ are point identified. To see this, note that in the special case of noninformative prior beliefs about $\mathbf{B}$ and $\mathbf{D}$, the posterior (17) simplifies to

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A})|\det(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')|^{T/2}}{\left\{\det\left[\mathrm{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')\right]\right\}^{T/2}} \tag{19}$$

where $\mathrm{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')$ denotes a matrix whose diagonal elements are the same as those of $\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}'$ and whose off-diagonal elements are zero. Thus when evaluated at any value of $\mathbf{A}$ that diagonalizes $\hat{\mathbf{\Omega}}_T$, the posterior distribution is proportional to the prior. Recall further from Hadamard's Inequality that if $\mathbf{A}$ has full rank and $\hat{\mathbf{\Omega}}_T$ is positive definite, then

$$\det\left[\mathrm{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')\right] \geq \det\left[\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}'\right]$$

with equality only if $\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}'$ is diagonal. Thus if we define

$$S(\mathbf{\Omega}) = \{\mathbf{\Omega} : \mathbf{A}\mathbf{\Omega}\mathbf{A}' = \mathrm{diag}(\mathbf{A}\mathbf{\Omega}\mathbf{A}')\}, \tag{20}$$

then

$$
\begin{aligned}
p(\mathbf{A}|\mathbf{Y}_T) &= k_T p(\mathbf{A}) && \text{if } \mathbf{A} \in S(\hat{\mathbf{\Omega}}_T) \\
&\xrightarrow{p} 0 && \text{if } \mathbf{A} \notin S(\hat{\mathbf{\Omega}}_T)
\end{aligned}
$$

More formally, for any $\mathbf{A}$ and $\mathbf{\Omega}$ we can measure the distance $q(\mathbf{A}, \mathbf{\Omega})$ between $\mathbf{A}$ and $S(\mathbf{\Omega})$ by the sum of squares of the off-diagonal elements of the Cholesky factor of $\mathbf{A\Omega A}'$,

$$
q(\mathbf{A}, \mathbf{\Omega}) = \textstyle\sum_{i=2}^{n} \sum_{j=1}^{i-1} p_{ij}^2(\mathbf{A}, \mathbf{\Omega}) \qquad \mathbf{P}(\mathbf{A}, \mathbf{\Omega})[\mathbf{P}(\mathbf{A}, \mathbf{\Omega})]' = \mathbf{A\Omega A}', \tag{21}
$$

so that $q(\mathbf{A}, \mathbf{\Omega}) = 0$ if and only if $\mathbf{A} \in S(\mathbf{\Omega})$. Let $H_\delta(\mathbf{\Omega})$ to be the set of all $\mathbf{A}$ that are within a distance δ of the set $S(\mathbf{\Omega})$:

$$
H_\delta(\mathbf{\Omega}) = \{\mathbf{A} : q(\mathbf{A}, \mathbf{\Omega}) \leq \delta\}. \tag{22}
$$

As long as the prior puts nonzero mass on some values of $\mathbf{A}$ that are consistent with the true $\mathbf{\Omega}_0$ (Prob$[\mathbf{A} \in H_\delta(\mathbf{\Omega}_0)] > 0$, $\forall \delta > 0$), then asymptotically the posterior will have no mass outside of this set (Prob$\{[\mathbf{A} \in H_\delta(\mathbf{\Omega}_0)]|\mathbf{Y}_T\} \to 1$, $\forall \delta > 0$). Proposition 2 summarizes the above asymptotic claims; see Appendix C for the proofs.

Proposition 2. *Let* $\underset{(n\times 1)}{\mathbf{y}_t}$ *be any process that is covariance stationary and ergodic for second moments. Let* $\underset{(k\times 1)}{\mathbf{x}_t} = (\mathbf{y}_t', \mathbf{y}_{t-1}', ..., \mathbf{y}_{t-m+1}', 1)'$ *and* $\mathbf{Y}_T = (\mathbf{x}_0', \mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_T')'$. *Define*

$$
\underset{(n\times k)}{\mathbf{\Phi}_0} = E(\mathbf{y}_t \mathbf{x}_{t-1}') \{E(\mathbf{x}_t \mathbf{x}_t')\}^{-1}
$$

$$
\underset{(n\times n)}{\mathbf{\Omega}_0} = E(\mathbf{y}_t \mathbf{x}_{t-1}') - E(\mathbf{y}_t \mathbf{x}_{t-1}') \{E(\mathbf{x}_t \mathbf{x}_t')\}^{-1} E(\mathbf{x}_{t-1} \mathbf{y}_t')
$$

with $E(\mathbf{x}_t\mathbf{x}_t')$ *and* $\mathbf{\Omega}_0$ *both assumed to be positive definite. Define* $\left\{\underset{(n\times n)}{\mathbf{A}}, d_{11}^{-1}, ..., d_{nn}^{-1}, \underset{(n\times k)}{\mathbf{B}}\right\}$ *to be random variables whose joint density conditional on* $\mathbf{Y}_T$ *is given by*

$$p(\mathbf{A}, d_{11}^{-1}, ..., d_{nn}^{-1}, \mathbf{B}|\mathbf{Y}_T) = k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T^*\mathbf{A}')]^{T/2} \times \prod_{i=1}^n \frac{\gamma(d_{ii}^{-1}; \kappa_i^*, \tau_i^*)\phi(\mathbf{b}_i; \mathbf{m}_i^*, d_{ii}\mathbf{M}_i^*)}{[(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i]^{\kappa_i+T/2}} \tag{23}$$

where $\mathbf{a}_i'$ *and* $\mathbf{b}_i'$ *denote the* i*th rows of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively, and* $\hat{\mathbf{\Omega}}_{iT}^*, \kappa_i^*, \tau_i^*, \mathbf{m}_i^*, \mathbf{M}_i^*$ *are the functions of* $\mathbf{Y}_T$ *defined in Proposition 1 where* $\mathbf{M}_i$ *is any invertible* $(k\times k)$ *matrix,* $\mathbf{m}_i' = \mathbf{a}_i'\boldsymbol{\eta}$ *for* $\boldsymbol{\eta}$ *any finite* $(n\times k)$ *matrix,* τ_i *and* κ_i *any finite nonnegative constants, and* k_T *an integrating constant that depends only on* $\hat{\mathbf{\Omega}}_{iT}^*, \tau_i, \kappa_i$ $(i = 1, ..., n)$ *and the functional form of* $p(\mathbf{A})$*. Let* $p(\mathbf{A})$ *be any bounded density for which* $\int_{\mathbf{A}\in H_\delta(\mathbf{\Omega}_0)} p(\mathbf{A})d\mathbf{A} > 0$ *for all* $\delta > 0$ *with* $H_\delta(\mathbf{\Omega})$ *defined in (22). Then as the sample size* T *goes to infinity, the random variables characterized by (23) have the following properties:*

(*i*) $\mathbf{B}|\mathbf{A}, d_{11}, ..., d_{nn}, \mathbf{Y}_T \xrightarrow{p} \mathbf{A}\mathbf{\Phi}_0$;

(*ii*) $\hat{\mathbf{\Omega}}_{iT}^* \xrightarrow{p} \mathbf{\Omega}_0$;

(*iii*) $d_{ii}|\mathbf{A}, \mathbf{Y}_T \xrightarrow{p} \mathbf{a}_i'\mathbf{\Omega}_0\mathbf{a}_i$;

(*iv*) $\text{Prob}[\mathbf{A} \in H_\delta(\mathbf{\Omega}_0)|\mathbf{Y}_T] \to 1$ for all $\delta > 0$.

Moreover, if $\kappa_i = \tau_i = 0$ and $\mathbf{M}_i = \mathbf{M}$ for $i = 1, ..., n$, *then when evaluated at any* $\mathbf{A} \in S(\hat{\mathbf{\Omega}}_T^*)$,

$$(v) \qquad p(\mathbf{A}|\mathbf{Y}_T) = k_T p(\mathbf{A}) \qquad \textit{for all } T.$$

Note that while we originally motivated the expression in (23) as the Bayesian posterior distribution for a Gaussian structural VAR, the results in Proposition 2 do not assume that the actual data are Gaussian or even that they follow a VAR. Nor does the proposition make any use of the fact that there exists a Bayesian interpretation of these formulas. The proposition provides a frequentist interpretation of what Bayesian inference amounts to when the sample gets large. The proposition establishes that as long as the prior assigns nonzero probability to a subset of $\mathbf{A}$ that diagonalizes the value $\mathbf{\Omega}_0$ defined in the proposition, then asymptotically the posterior density will be confined to that subset and at any point within the set will converge to some constant times the value of the prior density at that point.

In the special case where the model is point-identified, there exists only one allowable value of $\mathbf{A}$ for which $\mathbf{A\Omega}_0\mathbf{A}'$ is diagonal. Provided that $p(\mathbf{A})$ is nonzero in a neighborhood including that point, the posterior distribution collapses to the Dirac delta function at this value of $\mathbf{A}$. This reproduces the familiar result that under point identification, the priors on all parameters ($\mathbf{A}, \mathbf{D}$, and $\mathbf{B}$) are asymptotically irrelevant and Bayesian inference is asymptotically equivalent to maximum likelihood estimation, producing consistent estimates of parameters.

For finite T, note that the posterior (17) reflects uncertainty about $\mathbf{A}$ that results not just from the fact that the data cannot distinguish between alternative values for $\mathbf{A}$ within the set $S(\mathbf{\Omega})$, but also uncertainty about the set $S(\mathbf{\Omega})$ itself due to sampling uncertainty associated with $\hat{\mathbf{\Omega}}_T$. Recall that the log likelihood for a model that imposes no restrictions at all on $\mathbf{A}$, $\mathbf{B}$, or $\mathbf{D}$ is given by

$$\mathcal{L}(\mathbf{Y}_T) = -(Tn/2)[1+\log(2\pi)] - (T/2)\log\left[\det(\hat{\mathbf{\Omega}}_T)\right]. \tag{24}$$

Comparing (24) with (19), it is clear with noninformative priors on $\mathbf{D}$ and $\mathbf{B}$, the height of the posterior at some specified $\mathbf{A}$ is a transformation of the likelihood ratio test of the hypothesis that the true $\mathbf{A}\boldsymbol{\Omega}\mathbf{A}'$ is diagonal. The Bayesian approach thus provides a nice tool for treating the two sources of uncertainty– uncertainty about alternative values within the set $S(\boldsymbol{\Omega})$, and uncertainty about the boundaries of the set $S(\boldsymbol{\Omega})$ itself– symmetrically, so that an optimal statistical or policy decision given the combined sources of uncertainty could be reached.

3 Sign restrictions for higher-horizon impacts.

In an effort to try to gain additional identification, many applied researchers impose sign restrictions not just on the time-zero structural impacts $\partial \mathbf{y}_t/\partial \mathbf{u}_t'$ but also on impacts $\partial \mathbf{y}_{t+s}/\partial \mathbf{u}_t'$ for some horizons $s = 0, 1, ..., S$. These are given by

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{u}_t'} = \boldsymbol{\Psi}_s \mathbf{A}^{-1} \tag{25}$$

for $\boldsymbol{\Psi}_s$ the first n rows and columns of $\mathbf{F}^s$ for

$$\mathbf{F} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_{m-1} & \boldsymbol{\Phi}_m \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix} \tag{26}$$

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_m \mathbf{y}_{t-m} + \boldsymbol{\varepsilon}_t.$$

In particular,

$$\frac{\partial \mathbf{y}_t}{\partial \mathbf{u}_t'} = \mathbf{A}^{-1} \tag{27}$$

$$\frac{\partial \mathbf{y}_{t+1}}{\partial \mathbf{u}_t'} = \mathbf{\Phi}_1 \mathbf{A}^{-1}. \tag{28}$$

Consider first the case in which $\mathbf{\Phi}_1$ is diagonal. If all the diagonal elements are positive, the signs of $\partial \mathbf{y}_{t+1}/\partial \mathbf{u}_t'$ are identical to those of $\mathbf{A}^{-1}$ itself. In this case, if $\mathbf{\Phi}_1$ were known, there would be zero additional information from the signs of $\partial \mathbf{y}_{t+1}/\partial \mathbf{u}_t'$ beyond that contained in the signs of $\partial \mathbf{y}_t/\partial \mathbf{u}_t'$. Alternatively, if $\mathbf{\Phi}_1$ is diagonal but the ith element is negative, the signs of $\partial y_{i,t+1}/\partial \mathbf{u}_t'$ are opposite those of $\partial y_{it}/\partial \mathbf{u}_t'$. In this case, as the sample size grows to infinity, there will be no posterior distribution satisfying a restriction such as $\partial y_{it}/\partial u_{jt}$ and $\partial y_{i,t+1}/\partial u_{jt}$ are both positive. In a finite sample, a simulated draw from the posterior distribution purporting to impose such a restriction would at best be purely an artifact of sampling error. Canova and Paustian (2011) demonstrated using a popular macro model that implications for the signs of structural multipliers beyond the zero horizon ($\partial \mathbf{y}_{t+s}/\partial \mathbf{u}_t'$ for $s > 0$) are generally not robust.

In the case when off-diagonal elements of $\mathbf{\Phi}_1$ are nonzero, there could be some differences in the signs of the effects of the shocks at horizons 0 and 1. However, it seems odd to view nonzero off-diagonal entries in $\mathbf{\Phi}_1$ as giving us additional information about the values of $\mathbf{A}$. Instead, it seems that what researchers primarily have in mind when using sign restrictions at multiple horizons is a prior belief that $\mathbf{\Phi}_1$ is similar to a matrix with positive diagonal entries and modest off-diagonal elements, in which case the effect of shocks at horizon 0 should be the same sign as the effect at horizon 1. However, this is a prior belief not about $\mathbf{A}$ but rather about the value of $\mathbf{\Phi}_1$. The appropriate way to represent this prior information from a Bayesian perspective is by specifying additional beliefs about $\mathbf{\Phi}_1, ..., \mathbf{\Phi}_m$, as opposed to refining our beliefs about the value of $\mathbf{A}$.

A prior expectation that $\mathbf{\Phi}_1 = \mathbf{I}_n$ and $\mathbf{\Phi}_2 = \mathbf{\Phi}_3 = \cdots = \mathbf{\Phi}_m = \mathbf{0}$ would correspond to a prior expectation that the structural shocks $\mathbf{u}_t$ have the same effect on $\mathbf{y}_{t+s}$ for all s. Nudging the unrestricted OLS estimates in the direction of such a prior has long been known to help improve the forecasting accuracy of a VAR.[5] This suggests that we might want to use priors for $\mathbf{A}$ and $\mathbf{B}$ that imply a value for $\boldsymbol{\eta} = E(\mathbf{\Phi})$ in Proposition 1 given by

$$\underset{(n\times k)}{\boldsymbol{\eta}} = \begin{bmatrix} \underset{(n\times n)}{\mathbf{I}_n} & \underset{[n\times(k-n)]}{\mathbf{0}} \end{bmatrix}. \tag{29}$$

As noted by Sims and Zha (1998), since $\mathbf{B} = \mathbf{A\Phi}$, this calls for setting the prior mean for $\mathbf{B}|\mathbf{A}$ to be

$$E(\mathbf{B}|\mathbf{A}) = \mathbf{A}\boldsymbol{\eta}$$

so $\mathbf{m}_i = E(\mathbf{b}_i|\mathbf{A}) = \boldsymbol{\eta}'\mathbf{a}_i$. We can also follow Doan, Litterman and Sims (1984) in putting more confidence in our prior beliefs that higher-order lags are zero, as we describe in detail in Appendix D.

Note however that while such a prior can improve estimates of the autoregressive coefficients $\mathbf{\Phi}_1, ..., \mathbf{\Phi}_m$ and reduced-form variance matrix $\mathbf{\Omega}$, it does not provide a separate basis for estimation of $\mathbf{A}$. The best it can do is help a finite sample more rapidly approach the limiting distribution in Proposition 2. Regardless of the strength one wants to place on prior beliefs about persistence of shocks, the data are uninformative for purposes of distinguishing between alternative elements within the set $S(\mathbf{\Omega})$.

[5] See for example Doan, Litterman and Sims (1984), Litterman (1986), and Smets and Wouters (2003).

4 Bayesian inference in a 2-variable dynamic market model.

To illustrate the above results, we consider in this section a simple model in which the observed variables $\mathbf{y}_t = (p_t, q_t)'$ are the logs of price and quantity and the structural equations are the supply and demand schedules:

supply: $q_t = k^s + \alpha^s p_t + b_{11}^s p_{t-1} + b_{12}^s q_{t-1} + b_{21}^s p_{t-2} + b_{22}^s q_{t-2} + \cdots + b_{m1}^s p_{t-m} + b_{m2}^s q_{t-m} + u_t^s$

$$\text{demand: } q_t = k^d + \beta^d p_t + b_{11}^d p_{t-1} + b_{12}^d q_{t-1} + b_{21}^d p_{t-2} + b_{22}^d q_{t-2} + \cdots + b_{m1}^d p_{t-m} + b_{m2}^d q_{t-m} + u_t^d. \quad (30)$$

Here α^s, the short-run price elasticity of supply, is presumed to be positive, while β^d, the short-run price elasticity of demand, should be negative. Note that the system (30) is a special case of (1) with

$$\mathbf{A} = \begin{bmatrix} -\alpha^s & 1 \\ -\beta^d & 1 \end{bmatrix}. \quad (31)$$

As in Shapiro and Watson (1988), for any given β one can find the maximum likelihood estimate of α by an IV regression of $\hat{\varepsilon}_{2t}$ on $\hat{\varepsilon}_{1t}$ using $\hat{\varepsilon}_{2t} - \beta\hat{\varepsilon}_{1t}$ as instruments, where $\hat{\varepsilon}_{it}$ are the residuals from OLS estimation of the reduced-form VAR:

$$\hat{\alpha}(\beta) = \frac{\sum_{t=1}^{T}(\hat{\varepsilon}_{2t} - \beta\hat{\varepsilon}_{1t})\hat{\varepsilon}_{2t}}{\sum_{t=1}^{T}(\hat{\varepsilon}_{2t} - \beta\hat{\varepsilon}_{1t})\hat{\varepsilon}_{1t}} = \frac{(\hat{\omega}_{22} - \beta\hat{\omega}_{12})}{(\hat{\omega}_{12} - \beta\hat{\omega}_{11})} \quad (32)$$

for $\hat{\omega}_{ij}$ the row i, column j element of $\hat{\mathbf{\Omega}} = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$. One can verify directly that any pair (α, β) satisfying (32) produces a diagonal matrix for $\mathbf{A}\hat{\mathbf{\Omega}}\mathbf{A}'$.

The top panel of Figure 1 shows what the function $\alpha(\beta)$ looks like for the case when the OLS residuals are negatively correlated. The particular values for $\hat{\mathbf{\Omega}}$ used to construct

Figure 1 are based on the estimated variance matrix of a VAR(4) for the growth rate of the U.S. CPI and real GDP with the regressions estimated for t = 1986:Q1 to 2008:Q3:[6]

$$\hat{\mathbf{\Omega}} = \begin{bmatrix} 0.1129 & -0.0308 \\ -0.0308 & 0.2114 \end{bmatrix}. \tag{33}$$

Thus any pair (α, β) lying on the black function plotted in the top panel of Figure 1 would maximize the likelihood function, and there is no basis in the data for preferring one point on the curve to any other.

If we restrict the supply elasticity α to be positive and the demand elasticity β to be negative, we are left with the upper left quadrant in the figure. When, as in this example, the OLS residuals are negatively correlated, the sign restriction is consistent with any $\alpha \in (0, \infty)$, but requires β to fall in (x_L, x_H) where

$$x_L = \hat{\omega}_{22}/\hat{\omega}_{12}$$

$$x_H = \hat{\omega}_{12}/\hat{\omega}_{11}.$$

Values of β outside this range are ruled out by the restriction that α must be positive. To see why this is true algebraically, note first that for $\beta < x_L$, the numerator of (32) becomes negative with the denominator positive.[7] Thus $\beta < x_L$ would imply a negative value for α which is ruled out. At the other end, when β rises above x_H, the value of α would switch from $+\infty$ to $-\infty$. Hence a positive α requires $\beta \in (x_L, x_H)$.

[6] This is the period associated with the Great Moderation. For most other subsamples of postwar U.S. data, the correlation between the residuals is smaller in absolute value, in which case the informativeness of sign restrictions is even less than for the case analyzed here, as we shall see below.

[7] The denominator of (32) is monotonically decreasing in β and at $\beta = x_L$ is equal to $(\hat{\omega}_{12}^2 - \hat{\omega}_{22}\hat{\omega}_{11})/\hat{\omega}_{12} > 0$ whenever $\hat{\omega}_{12} < 0$ and $\hat{\mathbf{\Omega}}$ is positive definite.

To see the intuition behind these bounds, note that x_H is the estimated coefficient from an OLS regression of $\hat{\varepsilon}_{2t}$ on $\hat{\varepsilon}_{1t}$, which is a weighted average of the positive supply elasticity α and negative demand elasticity β (see for example Hamilton, 1994, equation [9.1.6]). Hence the MLE for β can be no larger than x_H and the MLE for α can be no smaller than x_H. The fact that the MLE for α can be no smaller than x_H is not a restriction, because we have separately required that $\alpha > 0$ and in the case under discussion, $x_H < 0$. However, the inference that β can be no larger than x_H has content. At the other end, the OLS coefficient from a regression of $\hat{\varepsilon}_{1t}$ on $\hat{\varepsilon}_{2t}$ (that is, x_L^{-1}) turns out to be a weighted average of α^{-1} and β^{-1}, requiring $\alpha^{-1} > x_L^{-1}$ (again not binding when $x_L < 0$) and $\beta^{-1} < x_L^{-1}$; the latter gives us the inference that $\beta > x_L$. This is the intuition for why $x_L < \beta < x_H$.

In the case when the correlation between the VAR residuals is instead positive, maximum likelihood estimation is consistent with any value for the demand elasticity β, while the supply elasticity α is constrained to fall in the interval (x_L, x_H), which when $\hat{\omega}_{12} > 0$ is a subset of the positive real line.

The bottom panel in Figure 1 plots contours of the concentrated likelihood function for U.S. inflation and output growth, that is, contours of the function

$$T \log |\det(\mathbf{A})| - (T/2) \log \left\{ \det \left[\mathrm{diag}(\mathbf{A}\hat{\mathbf{\Omega}}\mathbf{A}') \right] \right\}.$$

The data are relatively informative that α and β should be close to the values that diagonalize $\hat{\mathbf{\Omega}}$, that is, that α and β are close to the function $\alpha = \alpha(\beta)$ shown in black, but the data give us no basis for choosing some values within this set over others.

The set $S(\mathbf{\Omega})$ in expression (20) is calculated for this example as follows. When the correlation between the VAR residuals ω_{12} is negative, $S(\mathbf{\Omega})$ is the set of all $\mathbf{A}$ in (31) such that $\alpha > 0$, $(\omega_{22}/\omega_{12}) < \beta < (\omega_{12}/\omega_{11})$, and $\alpha = (\omega_{22} - \beta\omega_{12})/(\omega_{12} - \beta\omega_{11})$, in other words, the set of points on the black curve in Figure 1 between x_L and x_H.

As an illustration of our suggested approach, we use noninformative priors for $\mathbf{D}$ and $\mathbf{B}$, so that the matrix $\hat{\mathbf{\Omega}}_T^*$ in (16) is simply the unrestricted VAR estimate $\hat{\mathbf{\Omega}}$ in (33). We represent prior beliefs about the elasticities α and β using truncated Student t distributions:

$$p(\alpha) = \begin{cases} [1 - F(0; c_\alpha, \sigma_\alpha, \nu_\alpha)]^{-1} f(\alpha; c_\alpha, \sigma_\alpha, \nu_\alpha) & \text{if } \alpha \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

$$p(\beta) = \begin{cases} F(0; c_\beta, \sigma_\beta, \nu_\beta)^{-1} f(\beta; c_\beta, \sigma_\beta, \nu_\beta) & \text{if } \beta \leq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{35}$$

Here $f(x; c, \sigma, \nu)$ denotes the density for a Student t variable with location c, scale σ, and degrees of freedom ν evaluated at x,

$$f(x; c, \sigma, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\sigma\Gamma(\nu/2)} \left(1 + \frac{(x-c)^2}{\sigma^2\nu}\right)^{-(\nu+1)/2} \tag{36}$$

and $F(.)$ the cumulative distribution function $F(x; c, \sigma, \nu) = \int_{-\infty}^{x} f(z; c, \sigma, \nu) dz$. Here $c_\alpha > 0$ and $c_\beta < 0$ denote the prior modes, and σ and ν govern confidence in the prior, with larger σ and smaller ν registering more prior uncertainty. Including ν as a separate parameter to govern tail behavior allows the prior to have infinite variance when $\nu = 2$ and even an undefined mean when $\nu = 1$. Suppose we have the prior belief that a supply elasticity of $+1$ and demand elasticity of -1 were the most likely values, with uncertainty around these numbers represented with a scale parameter of unity and 5 degrees of freedom:

$$c_\alpha = 1, c_\beta = -1, \sigma_\alpha = \sigma_\beta = 1, \nu_\alpha = \nu_\beta = 5.$$

Contours for this prior distribution are provided in the top panel of Figure 2. A researcher using these priors would be relatively surprised to see an elasticity greater than 2 in absolute value. The bottom panel displays contours for the posterior distribution. The prior distribution leads the researcher to favor those points along the $\alpha(\beta)$ schedule associated with values for the supply and demand elasticities that are less than 2 in absolute value. If the sample size were to grow to infinity, the posterior would simply become the prior distribution constrained to the set $S(\mathbf{\Omega})$.

We plot the prior and posterior marginal distributions for α and β in Figure 3. Even though the data do not restrict the value of α, it turns out they are informative about α in a Bayesian sense given our prior beliefs. Prior to seeing the data, we thought very low values for the supply elasticity were fairly likely. However, given the data, a value for the supply elasticity below 0.5 would require a demand elasticity below -2.5 (see Figure 1), the latter having been regarded as a priori unlikely. For this reason, the posterior has less mass on the lower supply elasticities than the prior, though the adjustment is not large. The data are a little more informative about the demand elasticity, since the data cast considerable doubt on a value of $\beta > x_H = -0.273$.

Our recommendation is that researchers routinely report the way in which the data add information (or fail to add information) relative to prior beliefs. In a more complicated model for which this cannot be calculated analytically as here, or if the researcher does not have time for these analytical calculations, we recommend that the researcher should characterize the posterior distribution by the numerical simulation algorithm given in Appendix B, and plot the histogram of the draws from the posterior distribution against the prior distribution from which the analysis derives. We will provide additional illustrations of this approach below.

4.1 Implicit prior and posterior distributions associated with the conventional sign-restriction methodology.

The above results might seem foreign to many users of sign-restricted VARs, because the algorithms typically adopted do not explicate the forms of priors such as $p(\alpha)$ or $p(\beta)$. In this subsection we demonstrate how the results of the preceding subsection relate to conventional methods for inference in sign-restricted VARs.

One difference is that whereas we have focused on estimation of the structural parameters in $\mathbf{A}$, most sign-restricted analyses instead pose the question in terms of the impacts of structural shocks at horizon 0,

$$\mathbf{H} = \frac{\partial \mathbf{y}_t}{\partial \mathbf{u}_t'},$$

which from (27) means $\mathbf{H} = \mathbf{A}^{-1}$. At least in the case of $n = 2$ variables, it is simple to find a mapping between the two representations, since we could normalize using the convention that demand and supply shocks are two kinds of events that each result in a 1% increase in price. This amounts to normalizing $\mathbf{H}$ as

$$\begin{bmatrix} \varepsilon_t^p \\ \varepsilon_t^q \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ h & g \end{bmatrix} \begin{bmatrix} u_t^s \\ u_t^d \end{bmatrix}. \tag{37}$$

For this normalization, it turns out that $h = \beta^d$, the short-run price-elasticity of demand, while $g = \alpha^s$, the short-run price-elasticity of supply,[8] meaning that for this normalization, the impact parameters are numerically identical to the structural elasticities. The prior restriction that a supply shock that raises price should lower the quantity ($h < 0$) is identical to the prior restriction that the demand curve slopes down ($\beta^d < 0$).

In the typical application of a sign-restricted VAR, the vector of structural disturbances $\mathbf{u}_t$ is instead normalized to have identity variance matrix and the relation between structural shocks $\mathbf{u}_t$ and the reduced-form disturbances $\boldsymbol{\varepsilon}_t$ is described as

$$\boldsymbol{\varepsilon}_t = \mathbf{PQu}_t \tag{38}$$

where $\mathbf{P}$ is estimated from the Cholesky factor of the reduced-form variance matrix,

$$\hat{\mathbf{P}}\hat{\mathbf{P}}' = \hat{\boldsymbol{\Omega}} = T^{-1} \textstyle\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' \tag{39}$$

and $\mathbf{Q}$ is an orthogonal matrix ($\mathbf{QQ}' = \mathbf{I}_n$) that is uniformly distributed in the sense of Haar (1933). Rubio-Ramírez, Waggoner, and Zha (2010) adapted the ideas in Stewart (1980) to propose a constructive algorithm to generate draws for $\mathbf{Q}$ from this distribution, which we briefly summarize in Appendix E.[9]

---

[8] Specifically, if we let $\mathbf{u}_t^*$ denote the vector of structural disturbances when the system is written in the form (30), then using (31), the relation $\boldsymbol{\varepsilon}_t = \mathbf{A}^{-1}\mathbf{u}_t^*$ can equivalently be written

$$\begin{aligned} \begin{bmatrix} \varepsilon_t^p \\ \varepsilon_t^q \end{bmatrix} &= (\beta^d - \alpha^s)^{-1} \begin{bmatrix} 1 & -1 \\ \beta^d & -\alpha^s \end{bmatrix} \begin{bmatrix} u_t^{*s} \\ u_t^{*d} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ \beta^d & \alpha^s \end{bmatrix} \begin{bmatrix} (\beta^d - \alpha^s)^{-1} u_t^{*s} \\ -(\beta^d - \alpha^s)^{-1} u_t^{*d} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ h & g \end{bmatrix} \begin{bmatrix} u_t^s \\ u_t^d \end{bmatrix} \end{aligned}$$

for $h = \beta^d$, $g = \alpha^s$, $u_t^s = (\beta^d - \alpha^s)^{-1} u_t^{*s}$, and $u_t^d = -(\beta^d - \alpha^s)^{-1} u_t^{*d}$.

[9] Arias, Rubio-Ramírez, and Waggoner (2013) have recently generalized these algorithms to handle both zero and sign restrictions.

The implicit prior associated with this procedure is best understood as a joint distribution $p(\boldsymbol{\Omega}, \mathbf{Q})$ which has been factored as the product of a marginal distribution for $\boldsymbol{\Omega}$ and a distribution for $\mathbf{Q}$ conditional on $\boldsymbol{\Omega}$, $p(\boldsymbol{\Omega}, \mathbf{Q}) = p(\boldsymbol{\Omega})p(\mathbf{Q}|\boldsymbol{\Omega})$. The prior for $\boldsymbol{\Omega}$ can be viewed as a Wishart distribution for the inverse of the reduced-form residual variance matrix,

$$
\begin{aligned}
p(\boldsymbol{\Omega}^{-1}) &= \left[2^{Nn/2}\pi^{n(n-1)/4}\prod_{j=1}^{n}\Gamma\left(\frac{N+1-j}{2}\right)\right]^{-1} \times \\
&\quad |\boldsymbol{\Lambda}|^{n/2}|\boldsymbol{\Omega}|^{-(N-n-1)/2}\exp\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\right)\right], \qquad (40)
\end{aligned}
$$

where N is the degrees of freedom and $\boldsymbol{\Lambda}$ is the scale matrix for the prior while n is the dimension of $\boldsymbol{\varepsilon}_t$. A diffuse prior for $\boldsymbol{\Omega}$ can be viewed as the limiting case as $N \to 0$ and $\boldsymbol{\Lambda} \to \mathbf{0}$. The prior for $\mathbf{Q}$ conditional on $\boldsymbol{\Omega}$ is a uniform Haar distribution among the set of orthogonal matrices truncated by the sign restrictions imposed on $\mathbf{PQ}$.

In the case when the number of observed variables $n = 2$, the set of orthogonal matrices $\mathbf{Q}$ can be characterized as either rotations[10]

$$
\mathbf{Q}_1 = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}
$$

or reflections

$$
\mathbf{Q}_2 = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}
$$

for some $\theta \in [-\pi, \pi]$. In this case, an algorithm for generating a draw from $\mathbf{Q}|\boldsymbol{\Omega}$ that is equivalent to that of Rubio-Ramírez, Waggoner, and Zha (2010) would be to first generate

[10] Caldara and Kamps (2012) use this rotation matrix to analyze the implications of the traditional sign-restriction methodology for the allowable range for $\mathbf{A}^{-1}$ in a particular numerical example.

θ from a uniform distribution over $[-\pi, \pi]$ and then select $\mathbf{Q}_1$ or $\mathbf{Q}_2$ with probability 1/2. The four parameters used in the previous subsection to describe the structural representation $\boldsymbol{\lambda} = (\alpha, \beta, \sigma_s^2, \sigma_d^2)'$ would for this method instead be summarized by $\boldsymbol{\lambda}^* = (\theta, p_{11}, p_{21}, p_{22})'$.

Using the rotation matrix $\mathbf{Q}_1$ for illustration, notice that

$$\begin{aligned} \mathbf{PQ}_1 &= \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} p_{11}\cos\theta & -p_{11}\sin\theta \\ (p_{21}\cos\theta + p_{22}\sin\theta) & (-p_{21}\sin\theta + p_{22}\cos\theta) \end{bmatrix}. \end{aligned} \tag{41}$$

Comparing (38) with (37), the relation between the parameterizations is

$$\beta^d = h = \frac{p_{21}\cos\theta + p_{22}\sin\theta}{p_{11}\cos\theta} = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}}\tan\theta \tag{42}$$

$$\alpha^s = g = \frac{p_{21}}{p_{11}} - \frac{p_{22}}{p_{11}}\cot\theta. \tag{43}$$

Recall that when $\theta \sim U(-\pi, \pi)$, $\tan\theta$ or $\cot\theta$ have a standard Cauchy distribution (e.g., Gubner, 2006, p. 194).[11] Thus in the absence of sign restrictions, the priors for h and g conditional on $\boldsymbol{\Omega}$ are both Cauchy variables with scale p_{22}/p_{11} and location p_{21}/p_{11}, with a perfect correlation between h and g characterized by (42) and (43).

Recalling that a Cauchy distribution is a Student t with location zero, unit scale, and one degree of freedom, the implicit prior associated with the conventional approach is analogous to equations (34)-(35) with $c = 0, \sigma = 1, \nu = 1$.

[11] Actually, Gubner shows this for $\theta \sim U(-\pi/2, \pi/2)$, but the tangent function simply replicates its values for $[-\pi/2, \pi/2]$ when evaluated over $[-\pi, \pi]$.

Note furthermore that when $p_{21} < 0$, g will be zero at the points θ_1 and $\theta_2 \in [-\pi, \pi]$ for which $\cot\theta = p_{21}/p_{22}$; see Figure 4. Nonnegative g then requires $\theta \in (\theta_1, 0)$ or $\theta \in (\theta_2, \pi)$. As θ ranges over these intervals, g can take on any value between 0 and ∞. But at the lower end of either, h would be given by

$$h_L = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}} \tan\theta_i = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}} \frac{p_{22}}{p_{21}} = \frac{p_{21}^2 + p_{22}^2}{p_{11}p_{21}} = \frac{\omega_{22}}{\omega_{12}}$$

with the last equality following from (39). At the other end, since $\tan(0) = \tan(\pi) = 0$, the largest value of h is characterized by

$$h_H = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}} \times 0 = \frac{p_{21}}{p_{11}} = \frac{p_{21}p_{11}}{p_{11}^2} = \frac{\omega_{12}}{\omega_{11}}.$$

These will be recognized as exactly the same values for x_L and x_H that were derived in the previous subsection. In other words, although the traditional methodology is not usually expressed in terms of an explicit prior and posterior distribution, it can be characterized in exactly those terms. When $\omega_{12} < 0$, the prior for g conditional on $\boldsymbol{\Omega}$ is Cauchy with location parameter $x_H = p_{21}/p_{11}$ and scale parameter p_{22}/p_{11}, truncated by $g > 0$, while the prior for h conditional on $\boldsymbol{\Omega}$ is Cauchy with those same location and scale parameters truncated by $x_L < h < x_H$.

We applied the traditional sign-restriction methodology (detailed in Appendix E) to the postwar data set on inflation and output described above. The simulated posterior distribution for the date zero impacts is plotted using solid bars, while the Cauchy priors just described are shown as smooth curves in Figure 5. These will be recognized as similar to some of the shapes found numerically for posterior distributions in a 5-variable VAR analyzed by Arias, Rubio-Ramírez, and Waggoner (2013), though these authors did not note the connection to the implicit prior nor provide the theory for why this is what would be found from the method.

Another implication of these analytical results is that it would generally be necessary to report the posterior medians rather than posterior means if the uniform Haar prior is used because the posterior mean for the impact of a demand shock does not exist. We can also see analytically the results from applying the traditional methodology when the correlation between the OLS residuals is zero. In this case the posterior distribution for the time-zero impact of either a supply or a demand shock is simply a Cauchy distribution centered at zero, truncated to have the correct sign, and scaled by the ratio of the standard deviation of quantity surprises to price surprises.

4.2 Posterior distributions for other objects.

Up to this point in the section we have assumed that the researcher is interested in magnitudes like the supply elasticity α. Suppose instead that we were interested in the parameter $\zeta_1 = \alpha/(\alpha - \beta)$, which represents the effect on equilibrium quantity of a horizontal shift of the demand curve by an amount $\Delta q_t = 1\%$. Note that the sign restrictions $\alpha \geq 0$ and $\beta \leq 0$ impose the bounds $0 \leq \zeta_1 \leq 1$; a shift of the demand curve by 1% will in equilibrium result in an increase in quantity of less than 1%. Thus whereas the asymptotic posterior distribution for α would range over $(0, \infty)$ if the correlation between the VAR residuals ε_t^p and ε_t^q is negative, the posterior distribution for ζ_1 is bounded by definition. However, this does not mean that the data are any more informative about ζ_1 than they were about α

or β. Given prior beliefs about α and β, the data will asymptotically restrict the posterior distribution of (α, β) within a certain region. But within that region, the posterior distribution for (α, β) is simply a constant times the prior. Likewise, the posterior distribution of ζ_1 is simply a constant times the prior distribution for ζ_1 that is implied by the prior for (α, β) over the restricted region. Inference about magnitudes like ζ_1 is essentially the case explored by Caldara and Kamps (2012) in their analysis of the multiplier effects of fiscal policy.

Another magnitude that researchers often report is the effect on price of a one-standard-deviation shock to u_t^s.[12] We can see from (41) that conditional on $\mathbf{\Omega}$, the prior for this magnitude that is implied by the uniform Haar prior is given by $\zeta_2 = p_{11} \cos\theta$ where p_{11} is the square root of the variance of ε_t^p and $\theta \sim U(-\pi, \pi)$. Here again the proposed magnitude of interest ζ_2 is bounded by definition, since the component of the variance of ε_t^p that is attributed to u_t^s cannot exceed the total variance of ε_t^p. Asymptotically the data will inform perfectly about p_{11} and restrict θ to lie in $A = \{(\theta_1, 0) \cup (\theta_2, \pi)\}$. But within A, the inference about ζ_2 is simply going to reflect prior beliefs about θ. The posterior distribution of θ will be completely flat throughout the set A.

This result helps shed light on the approach suggested by Inoue and Kilian (2013). They proposed using the posterior mode of a magnitude like ζ_2 as the point estimate to be reported. Note that even though the posterior for θ is completely flat, the posterior for $\cos\theta$ has a

[12] Fry and Pagan (2011, p. 955) discussed why we might have more hope of saying something meaningful about the effects of one-standard-deviation shocks than about the supply and demand elasticities α and β, but also explained why the effect of a one-standard-deviation shock is usually not the magnitude we would be interested in.

unique maximum. However, given $\mathbf{\Omega}$, the location of this maximum, and indeed the shape of the posterior over its range, is determined by the shape of the prior. Although the prior may seem uninformative about θ, the same prior has a clear preference for values of ζ_2 near p_{11}. Our suggestion is that rather than pretend that our priors have had no influence on the reported results, it would be better for researchers to defend their prior beliefs explicitly, and clarify how those prior beliefs have ended up influencing the conclusions. We illustrate how this can be done in the following section.

5 Bayesian inference in a 3-variable macro model.

5.1 Model description.

Here we illustrate these methods using a commonly studied three-variable macroeconomic model.[13] 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). The system consists of a Phillips Curve,

$$y_t = k^s + \alpha^s \pi_t + [\mathbf{b}^s]' \mathbf{x}_{t-1} + u_t^s, \tag{44}$$

[13] Equations (44)-(46) can be motivated from the 3-variable macro models studied by Rotemberg and Woodford (1997), Del Negro and Schorfheide (2004), Giordani (2004), Benati and Surico (2009), and Rubio-Ramirez, Waggoner, and Zha (2010). Appendix F details the relation between our parameterization and that in Benati and Surico (2009).

an aggregate demand equation,

$$y_t = k^d + \beta^d \pi_t + \gamma^d r_t + \left[\mathbf{b}^d\right]' \mathbf{x}_{t-1} + u_t^d, \tag{45}$$

and a Taylor rule for monetary policy,

$$r_t = k^m + \psi^y y_t + \psi^\pi \pi_t + \left[\mathbf{b}^m\right]' \mathbf{x}_{t-1} + u_t^m, \tag{46}$$

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. This system will be recognized as a special case of the general framework (1) studied in Section 2 with

$$\mathbf{A} = \begin{bmatrix} 1 & -\alpha^s & 0 \\ 1 & -\beta^d & -\gamma^d \\ -\psi^y & -\psi^\pi & 1 \end{bmatrix}. \tag{47}$$

Many macroeconomists have strong prior beliefs about the values of key parameters in a system like (44)-(46), and indeed many studies simply assume particular numerical values for purposes of quantitative analysis. The modes of the prior distributions used in our study for elements of the contemporaneous coefficients in $\mathbf{A}$ (that is, the values for c in equation (36)) are summarized in Table 1. For the coefficients that characterize the Fed's response to the output gap (ψ^y) and inflation (ψ^π) our prior modes correspond to the values from Taylor's (1993) original article (0.5 and 1.5, respectively). We further impose the restriction that both these Taylor Rule parameters must be nonnegative. For the slope of the aggregate supply curve, our prior for α^s has mode at 2, consistent with the prior for this parameter assumed by Lubik and Schorfheide (2004). Our parameterization for the aggregate demand curve can be viewed as the reduced form of a forward-looking DSGE model in which aggregate demand responds negatively to the ex ante real interest rate. This motivated our choice of a prior mode of 0.75 for β^d and -1 for γ^d.[14] For each of the five contemporaneous coefficients, we set the scale parameter $\sigma = 0.3$ and the degrees of freedom $\nu = 2$. The priors thus have fat tails with infinite variance but finite mean, with much of the mass in a neighborhood around $c \pm 0.6$. These prior densities are plotted as the red curves in Figure 6.

Our priors for the variances of the structural shocks are based on the mode (denoted $\mathbf{A}^*$) of our prior distribution for $\mathbf{A}$ combined with the scales of individual innovations as summarized by univariate autoregressions fit to each individual series. Specifically, if $\hat{e}_{it}$ denotes 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 κ_i/τ_i (the prior mean for d_{ii}^{-1}) equal to the reciprocal of the ith diagonal element of $\mathbf{A}^*\mathbf{S}\mathbf{A}^{*\prime}$. We put only modest weight on these prior beliefs by setting each $\kappa_i = 2$.[15]

Our choice of priors on the lagged coefficients (summarized in equation (55)) are similar to the values suggested by Doan (2013). We set $\lambda_1 = 1$ (which governs how quickly the prior for lagged coefficients tightens to zero as the lag ℓ increases) , $\lambda_2 = 0.7$ (which governs

[14] If the reduced form for output and inflation behave approximately like univariate AR(1) processes with autoregressive coefficient $\phi = 0.75$, then $d_y = g_\pi = \phi$ and all other parameters in (60) and (61) would be zero. This leads to a prior belief from equation (56) that β^d would be around $\phi\sigma^{-1}/(1-\phi\eta)$ and γ^d around $-\sigma^{-1}/(1-\phi\eta)$. Benati and Surico (2009) used a prior mode of 0.25 for η (which is represented by γ in their equation (3)) while Lubik and Schorfheide (2004) in their equation (1) assumed $\eta = 1$. Both these studies (and most others) have taken the σ (the intertemporal elasticity of substitution) to be 2. Using $\eta = 2/3, \phi = 0.75$, and $\sigma = 2$ gives predicted values for γ^d of -1 and β^d of 0.75.

[15] Note from equation (13) that our prior is thus given the same weight as 4 observations out of a sample of size $T = 91$.

how much we shrink coefficients other than own lags to zero), and $\lambda_3 = 100$ (which makes the prior on the constant term essentially irrelevant). For our base case we set λ_0, the parameter controlling the overall tightness of the prior, to 0.2.

One way to visualize these priors is to consider what they imply for the impact of a given structural shock at time t on the value of the system at date $t + s$ as in (25),

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{u}_t'} = \boldsymbol{\Psi}_s \mathbf{A}^{-1}. \tag{48}$$

The traditional approach would impose the restriction that the impacts of structural shocks have particular signs for a subset of horizons $s = 0, 1, 2...$ By contrast, our approach calls on the researcher to specify priors for $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{B}$ and see what these imply for the likely impact of structural shocks at any horizon. Our priors imply a high probability that the initial impacts have the expected signs, but since this probability is less than one, also allow the data to override our prior beliefs for those objects about which the data are informative.

The top panel of Table 2 summarizes the probability implied by our priors that the impacts of the structural shocks given in (48) are initially positive ($s = 0$). A demand shock is viewed by our priors as extremely likely to increase output, inflation, and the interest rate, and a contractionary monetary shock to increase the interest rate and reduce output and inflation. A favorable supply shock almost surely decreases the current inflation and interest rate and with high probability is also associated with higher output.

Table 2 also reports the weight that our prior puts on the signs of structural impacts at horizons 1 and 2. The rate at which probabilities decay towards 0.5 as we increase s is governed by the tightness of the prior for the persistence parameters λ_k associated with the

Minnesota prior along with prior beliefs about the size of structural shocks as summarized by (κ_i, τ_i). As seen in Table 2, our use of a relatively tight Minnesota prior ($\lambda_0 = 0.2$) implies almost as much confidence in the signs of $\partial \mathbf{y}_{t+1}/\partial \mathbf{u}_t'$ and $\partial \mathbf{y}_{t+2}/\partial \mathbf{u}_t'$ as for $\partial \mathbf{y}_t/\partial \mathbf{u}_t'$.

5.2 Empirical results.

We used quarterly data on $\mathbf{y}_t$ with regressions estimated over $t =$ 1986:Q1 to 2008:Q3 to calculate posterior distributions for parameters using the algorithm described in Appendix B. Posterior distributions for the 5 contemporaneous coefficients are plotted as solid histograms in Figure 6. The data turn out to be highly informative about the values of ψ^π and β^d, cause modest revisions in our beliefs about ψ^y, and are relatively uninformative about α^s and γ^d. To understand what drives these results, notice that for any given values of ψ^y and ψ^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 $(\hat{\varepsilon}_t^r - \psi^y \hat{\varepsilon}_t^y - \psi^\pi \hat{\varepsilon}_t^\pi)$ as instrument:

$$\hat{\alpha}(\psi^y, \psi^\pi) = \frac{\hat{\omega}_{13} - \psi^y \hat{\omega}_{11} - \psi^\pi \hat{\omega}_{12}}{\hat{\omega}_{23} - \psi^y \hat{\omega}_{21} - \psi^\pi \hat{\omega}_{22}} \tag{49}$$

for $\hat{\omega}_{ij}$ the row i, column j element of

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

Figure 7 plots possible values for ψ^y on the horizontal axis and ψ^π on the vertical axis. The dashed line identifies combinations of ψ^y and ψ^π for which the numerator of (49) would be zero, that is, plots the line $\hat{\omega}_{13} - \psi^y \hat{\omega}_{11} - \psi^\pi \hat{\omega}_{12} = 0$. This line could alternatively be described as combinations of ψ^y and ψ^π for which the structural monetary shock ($u_t^m = \varepsilon_t^r - \psi^y \varepsilon_t^y - \psi^\pi \varepsilon_t^\pi$) 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 beaded line plots combinations of ψ^y and ψ^π for which the denominator of (49) is zero, namely $\hat{\omega}_{23} - \psi^y \hat{\omega}_{21} - \psi^\pi \hat{\omega}_{22} = 0$. 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 (49) would have to be of the same sign. Since $\hat{\omega}_{12} < 0$, the numerator is positive for any pair (ψ^y, ψ^π) above the dashed line and negative for any point below the dashed line. The denominator of (49) is positive for any point below the beaded line. Thus in order to satisfy $\alpha^s > 0$, the values of ψ^y and ψ^π would have to be in a shaded region of Figure 7, either both above the dashed line and below the beaded (the lower left quadrant of Figure 7) or below the dashed and above the beaded (the upper right quadrant). As one moves from the dashed line to the beaded line within a shaded region, the MLE of α^s would vary from 0 to $+\infty$. Thus if we had only the sign restrictions on ψ^y, ψ^π, and α^s, the data would rule out combinations for which one of the ψ's is small and the other large but put no restrictions on α^s.

The data's informativeness about β^d comes from interaction with the priors. Given the maximum likelihood estimate $\hat{\alpha}(\psi^y, \psi^\pi)$ associated with any specified (ψ^y, ψ^π), we can find the MLE for β^d and γ^d associated with that (ψ^y, ψ^π) by an IV regression of $\hat{\varepsilon}_t^y$ on $\hat{\varepsilon}_t^\pi$ and $\hat{\varepsilon}_t^r$

using $(\hat{\varepsilon}_t^y - \hat{\alpha}(\psi^y, \psi^\pi))$ and $(\hat{\varepsilon}_t^r - \psi^y \hat{\varepsilon}_t^y - \psi^\pi \hat{\varepsilon}_t^\pi)$ as instruments:

$$\begin{bmatrix} \hat{\beta}(\psi^y, \psi^\pi) \\ \\ \hat{\gamma}(\psi^y, \psi^\pi) \end{bmatrix} = \begin{bmatrix} [\hat{\omega}_{12} - \hat{\alpha}(\psi^y, \psi^\pi)\hat{\omega}_{22}] & [\hat{\omega}_{13} - \hat{\alpha}(\psi^y, \psi^\pi)\hat{\omega}_{23}] \\ \\ [\hat{\omega}_{32} - \psi^y \hat{\omega}_{12} - \psi^\pi \hat{\omega}_{22}] & [\hat{\omega}_{33} - \psi^y \hat{\omega}_{13} - \psi^\pi \hat{\omega}_{23}] \end{bmatrix}^{-1} \times \begin{bmatrix} [\hat{\omega}_{11} - \hat{\alpha}(\psi^y, \psi^\pi)\hat{\omega}_{12}] \\ \\ [\hat{\omega}_{31} - \psi^y \hat{\omega}_{11} - \psi^\pi \hat{\omega}_{21}] \end{bmatrix}. \tag{50}$$

One can use these expressions to calculate combinations of ψ^y and ψ^π that would imply a negative value for β^d. This corresponds to the region below the hatched curve and above the solid curve in Figure 7, 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 (50) 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 hatched (the upper bound). It turns out that there are very few values of ψ^y and ψ^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 ψ^y and ψ^π would consist of the light shaded areas in Figure 7. 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, ψ^y, and ψ^π that these parameter values would imply. Since these are regarded as also unlikely a priori, our posterior inference ends up putting considerable posterior probability on the dark shaded regions in Figure 7 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.

Table 2 summarizes the posterior probabilities that a specified structural shock results in an increase in the indicated variable at horizons $s = 0, 1, 2$. The posterior distribution is more tightly adhering to the expected signs of these impacts than was our prior distribution. A favorable supply shock is virtually certain to raise output and lower inflation, while a demand shock is virtually certain to raise all three variables. A monetary contraction is extremely likely to raise interest rates and lower output and inflation on impact, though we would have less than 95% confidence that the negative effect on output would last as long as two quarters.

Impulse-response functions are plotted in Figure 8. The solid lines plot the median of the posterior distribution for any given horizon. This would correspond to the optimal inference about the impact at that horizon if the researcher's loss function is based on absolute value. 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 8 represent 95% posterior credibility regions.

The first two columns of Figure 8 summarize the effects of supply and demand shocks, both of which are quite persistent in our baseline simulation with posterior confidence about the signs of effects lasting beyond a year. The third column in Figure 8 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, and these feed back into the interest rate. The Taylor Rule equation shifts up by 100 basis points, but within the month the economy moves along the new Taylor Rule equation with output falling 0.38% and inflation falling about 0.25%, as a result of which in equilibrium the fed funds rate is only about 65 basis points higher in the immediate response to the shock. The output effect declines relatively quickly in the quarters following the shock, with the point estimate actually switching signs after 5 quarters.

We also can explore the importance of prior beliefs about the persistence of shocks by relaxing the tightness of the Minnesota prior. We can represent the complete absence of information about the persistence of shocks by setting the overall tightness parameter $\lambda_0 = 10^9$. Impulse-response functions for this alternative specification of prior beliefs are plotted in Figure 9. In the absence of prior information about the persistence of shocks, the confidence intervals associated with the impulse-response functions become somewhat wider, though the overall conclusions are very similar to those for our baseline specification. This is consistent with the conclusion of many users of the conventional sign restriction methodology– prior beliefs about the effect of shocks at longer horizons can help improve the quality of overall inference. However, representing this information in the form suggested here as smooth distributions over $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{B}$ offers a number of advantages over the conventional sign-restriction approach.

6 Conclusion.

One reason that Bayesian methods are not more widely used in economics is the desire to let the data speak for themselves, uncolored by subjective or prior beliefs. Unfortunately, that laudable goal is not really achievable– some reliance on prior understanding is necessary in order to make any structural interpretation of observed data. The way this is done in most studies is to insist that we know with certainty some of the parameters about which a reasonable person would entertain some doubt, while claiming to know nothing at all a priori about other details about which we in fact have at least some information. Moreover, the process by which the imposed restrictions are arrived at in practice has often amounted to calculating impulse-response functions such as those in Figure 8 and verifying that they "look reasonable."

The methods proposed here provide a formal device for making use of prior information which has always been implicit, although often not acknowledged, in earlier efforts to draw structural inference from vector autoregressions. Explicitly specifying prior beliefs and then reporting the relation between prior and posterior distributions gives researchers a device with which to summarize more clearly those features of interest for which the data are informative and those for which they are not.

Appendix

A. Proof of Proposition 1.

The likelihood (10) can be written

$$p(\mathbf{Y}_T|\mathbf{A},\mathbf{D},\mathbf{B}) = (2\pi)^{-Tn/2}|\det(\mathbf{A})|^T|\mathbf{D}|^{-T/2}\prod\nolimits_{i=1}^{n}\exp\left[-\sum_{t=1}^{T}\frac{(\mathbf{a}_i'\mathbf{y}_t-\mathbf{b}_i'\mathbf{x}_{t-1})^2}{2d_{ii}}\right].$$

If we define $\mathbf{P}_i$ to be the Cholesky factor of $\mathbf{M}_i^{-1} = \mathbf{P}_i\mathbf{P}_i'$, $\mathbf{X}_i^* = \mathbf{P}_i$, and $\mathbf{y}_i^* = \mathbf{P}_i\mathbf{m}_i$, the prior for $\mathbf{b}_i$ in (8) can be written

$$p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) = \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}}\exp\left[-\frac{(\mathbf{y}_i^*-\mathbf{X}_i^*\mathbf{b}_i)'(\mathbf{y}_i^*-\mathbf{X}_i^*\mathbf{b}_i)}{2d_{ii}}\right].$$

Comparing the above two equations we see that, conditional on $\mathbf{A}$, prior information about $\mathbf{b}_i$ can be combined with the information in the data by regressing $\tilde{\mathbf{Y}}_i$ on $\tilde{\mathbf{X}}_i$:

$$\underset{[(T+k)\times 1]}{\tilde{\mathbf{Y}}_i} = \begin{bmatrix} \mathbf{y}_1'\mathbf{a}_i \\ \vdots \\ \mathbf{y}_T'\mathbf{a}_i \\ \mathbf{P}_i\mathbf{m}_i \end{bmatrix} \qquad \underset{[(T+k)\times k]}{\tilde{\mathbf{X}}_i} = \begin{bmatrix} \mathbf{x}_0' \\ \vdots \\ \mathbf{x}_{T-1}' \\ \mathbf{P}_i \end{bmatrix}$$

$$\begin{aligned} \mathbf{m}_i^* &= \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i \qquad (51)\\ &= \left(\sum\nolimits_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}' + \mathbf{M}_i^{-1}\right)^{-1}\left(\sum\nolimits_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t' + \mathbf{M}_i^{-1}\boldsymbol{\eta}'\right)\mathbf{a}_i. \end{aligned}$$

From the property that the OLS residuals $\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^*$ are orthogonal to $\tilde{\mathbf{X}}_i$, we further know

$$\begin{aligned}
(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{b}_i)'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{b}_i) &= (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^* + \tilde{\mathbf{X}}_i\mathbf{m}_i^* - \tilde{\mathbf{X}}_i\mathbf{b}_i)'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^* + \tilde{\mathbf{X}}_i\mathbf{m}_i^* - \tilde{\mathbf{X}}_i\mathbf{b}_i) \\
&= (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^*)'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^*) + (\mathbf{b}_i - \mathbf{m}_i^*)'\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i(\mathbf{b}_i - \mathbf{m}_i^*) \\
&= \tilde{\mathbf{Y}}_i'\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}_i'\tilde{\mathbf{X}}_i(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i + (\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*) \\
&= T\mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i + (\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*).
\end{aligned}$$

The product of the likelihood (10) with the prior for $\mathbf{B}$ (7) can thus be written

$$p(\mathbf{B}|\mathbf{A},\mathbf{D})p(\mathbf{Y}_T|\mathbf{A},\mathbf{D},\mathbf{B}) = (2\pi)^{-Tn/2}|\det(\mathbf{A})|^T|\mathbf{D}|^{-T/2} \times \tag{52}$$
$$\prod_{i=1}^n \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}} \exp\left[-\frac{T\mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i + (\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*)}{2d_{ii}}\right].$$

Multiplying (52) by the priors for $\mathbf{A}$ and $\mathbf{D}$ and rearranging gives

$$\begin{aligned}
p(\mathbf{Y}_T,\mathbf{A},\mathbf{D},\mathbf{B}) &= p(\mathbf{A})p(\mathbf{D}|\mathbf{A})p(\mathbf{B}|\mathbf{A},\mathbf{D})p(\mathbf{Y}_T|\mathbf{A},\mathbf{D},\mathbf{B}) \\
&= p(\mathbf{A})(2\pi)^{-Tn/2}|\det(\mathbf{A})|^T \prod_{i=1}^n \left\{ d_{ii}^{-T/2}\frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}\frac{\Gamma(\kappa_i^*)}{(\tau_i^*)^{\kappa_i^*}}\frac{(\tau_i^*)^{\kappa_i^*}}{\Gamma(\kappa_i^*)}(d_{ii}^{-1})^{\kappa_i-1}\exp(-\tau_i^* d_{ii}^{-1})\times \right. \\
&\qquad \left. \frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}}\frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i^*|^{1/2}}\exp\left[-\frac{(\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*)}{2d_{ii}}\right]\right\} \\
&= p(\mathbf{A})(2\pi)^{-Tn/2}|\det(\mathbf{A})|^T \times \\
&\qquad \prod_{i=1}^n \left\{\frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}}\frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}\frac{\Gamma(\kappa_i^*)}{(\tau_i^*)^{\kappa_i^*}}\right\}\gamma(d_{ii}^{-1};\kappa_i^*,\tau_i^*)\phi(\mathbf{b}_i;\mathbf{m}_i^*,d_{ii}\mathbf{M}_i^*). \qquad (53)
\end{aligned}$$

Note that the product in (53) can be interpreted as

$$p(\mathbf{Y}_T,\mathbf{A},\mathbf{D},\mathbf{B}) = p(\mathbf{Y}_T)p(\mathbf{A}|\mathbf{Y}_T)p(\mathbf{D}|\mathbf{A},\mathbf{Y}_T)p(\mathbf{B}|\mathbf{A},\mathbf{D},\mathbf{Y}_T).$$

Thus the posterior $p(\mathbf{B}|\mathbf{A},\mathbf{D},\mathbf{Y}_T)$ is the product of $N(\mathbf{m}_i^*, d_{ii}\mathbf{M}_i^*)$ densities, the posterior

$p(\mathbf{D}|\mathbf{A}, \mathbf{Y}_T)$ the product of $\Gamma(\kappa_i^*, \tau_i^*)$ densities, and

$$p(\mathbf{Y}_T)p(\mathbf{A}|\mathbf{Y}_T) = p(\mathbf{A})(2\pi)^{-Tn/2}|\det(\mathbf{A})|^T \prod_{i=1}^{n} \left\{ \frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)} \frac{\Gamma(\kappa_i^*)}{(\tau_i^*)^{\kappa_i^*}} \right\}$$

from which

$$p(\mathbf{A}|\mathbf{Y}_T) \propto \frac{p(\mathbf{A})|\det(\mathbf{A})|^T}{\prod_{i=1}^{n}(\tau_i + T\mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i/2)^{\kappa_i+T/2}}.$$

Since $\hat{\mathbf{\Omega}}_T^*$ is not a function of $\mathbf{A}$, we can write the above result in an equivalent form to facilitate numerical calculation and interpretation:

$$p(\mathbf{A}|\mathbf{Y}_T) \propto \frac{p(\mathbf{A})|\det(\mathbf{A})|^T|\hat{\mathbf{\Omega}}_T^*|^{T/2}}{\prod_{i=1}^{n}[(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i]^{\kappa_i+T/2}} = \frac{p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T^*\mathbf{A}')]^{T/2}}{\prod_{i=1}^{n}[(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i]^{\kappa_i+T/2}}$$

as claimed in equation (17). Note that we could replace $\hat{\mathbf{\Omega}}_T^*$ in the numerator with $\hat{\mathbf{\Omega}}_T$, $\mathbf{I}_n$, or any matrix not depending on unknown parameters, with any such replacement simply changing the definition of k_T in (17). Our use of $\hat{\mathbf{\Omega}}_T^*$ in the numerator helps the target density (which omits k_T) behave better numerically for large T, as will be seen in the asymptotic analysis below.

B. Numerical algorithm for drawing from the posterior distribution in Proposition 1.[16]

We use a random-walk Metropolis Hastings algorithm to draw from the posterior distribution of $\mathbf{A}$ and use the known closed-form expressions to generate draws from $\mathbf{D}|\mathbf{A}, \mathbf{Y}_T$ and $\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T$. We first calculate $\hat{\mathbf{\Omega}}_{iT}^*$ from (15) and

$$\mathbf{M}_i^* = \left(\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}_{t-1}' + \mathbf{M}_i^{-1}\right)^{-1}$$

[16] Code to implement this procedure is available at http://econweb.ucsd.edu/~jhamilton/BHcode.zip.

$$\hat{\mathbf{\Phi}}_i^* = \mathbf{M}_i^* \left(\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{y}_t' + \mathbf{M}_i^{-1}\boldsymbol{\eta}' \right)$$

for $i = 1, ..., n$ and $\boldsymbol{\eta}$ the prior beliefs as specified for example in (29). Note that these depend solely on the data and fixed parameters, so only need be calculated once prior to any iterations. Define the target function to be

$$\tilde{p}(\mathbf{A}) = \log p(\mathbf{A}) + (T/2)\log[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T^*\mathbf{A}')] - \sum_{i=1}^n (\kappa_i + T/2)\log[(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i].$$

Start with an arbitrary initial value for $\mathbf{A}^{(0)}$, drawn for example from the prior $p(\mathbf{A})$. Next, for $\ell = 1, ..., D$, we generate a candidate draw (denoted $\tilde{\mathbf{A}}^{(\ell)}$) from a proposal density $q(\mathbf{A}|\mathbf{A}^{(\ell-1)})$; in our applications we used a random-walk Metropolis Hastings algorithm:

$$q(\mathbf{A}|\mathbf{A}^{(\ell-1)}) = \phi(\text{vec}^*(\mathbf{A});\text{vec}^*(\mathbf{A}^{(\ell-1)}), \xi\mathbf{I}_c).$$

Here $\text{vec}^*(\mathbf{A})$ denotes the vector consisting of the unknown elements of $\mathbf{A}$, c the number of unknown elements, and $\phi(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Omega})$ indicates the multivariate Normal density with mean $\boldsymbol{\mu}$ and variance $\mathbf{\Omega}$ evaluated at $\mathbf{x}$, while ξ is a tuning parameter to achieve a target acceptance rate of around 35%. We then accept the candidate draw for $\mathbf{A}$ (setting $\mathbf{A}^{(\ell)} = \tilde{\mathbf{A}}^{(\ell)}$) with probability

$$\min\left\{\exp\left[\tilde{p}(\tilde{\mathbf{A}}^{(\ell)}) - \tilde{p}(\mathbf{A}^{(\ell-1)})\right], 1\right\}.$$

If we fail to accept the draw, we set $\mathbf{A}^{(\ell)} = \mathbf{A}^{(\ell-1)}$.

We then generate a draw for $\left[d_{ii}^{(\ell)}\right]^{-1}$ from independent $\Gamma(\kappa_i + T/2, \tau_i + T\mathbf{a}_i^{(\ell)\prime}\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i^{(\ell)}/2)$ densities and $\mathbf{b}_i$ from $N(\hat{\mathbf{\Phi}}_i^*\mathbf{a}_i^{(\ell)}, d_{ii}^{(\ell)}\mathbf{M}_i^*)$ densities for $i = 1, ..., n$. The triplet $\{\mathbf{A}^{(\ell)}, \mathbf{D}^{(\ell)}, \mathbf{B}^{(\ell)}\}$ represents a single draw from $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$ after we throw out the first 1,000,000 burn-in draws. The figures in the text are based on 1,000,000 post burn-in draws.

C. Proof of Proposition 2.

$$
\begin{aligned}
(i) \quad E[(\mathbf{b}_i-\mathbf{m}_i^*)(\mathbf{b}_i-\mathbf{m}_i^*)'|\mathbf{A},d_{11},...,d_{nn},\mathbf{Y}_T] &= \left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'+\mathbf{M}_i^{-1}\right)^{-1} \\
&= T^{-1}\left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'+T^{-1}\mathbf{M}_i^{-1}\right)^{-1} \\
&\xrightarrow{p}\mathbf{0}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{m}_i^* &= \left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'+\mathbf{M}_i^{-1}\right)^{-1}\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t'+\mathbf{M}_i^{-1}\boldsymbol{\eta}\right)\mathbf{a}_i \\
&= \left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'+T^{-1}\mathbf{M}_i^{-1}\right)^{-1}\left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t'+T^{-1}\mathbf{M}_i^{-1}\boldsymbol{\eta}\right)\mathbf{a}_i \\
&\xrightarrow{p}\boldsymbol{\Phi}_0'\mathbf{a}_i.
\end{aligned}
$$

Hence

$$
\mathbf{B}=\begin{bmatrix}\mathbf{b}_1' \\ \vdots \\ \mathbf{b}_n'\end{bmatrix}\xrightarrow{p}\mathbf{A}\boldsymbol{\Phi}_0.
$$

$$
\begin{aligned}
(ii) \quad \hat{\boldsymbol{\Omega}}_{iT}^* &= T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'+T^{-1}\boldsymbol{\eta}\mathbf{M}_i^{-1}\boldsymbol{\eta}'-\left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}'+T^{-1}\boldsymbol{\eta}\mathbf{M}_i^{-1}\right)\times \\
&\quad \left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'+T^{-1}\mathbf{M}_i^{-1}\right)^{-1}\left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t'+T^{-1}\mathbf{M}_i^{-1}\boldsymbol{\eta}'\right) \\
&\quad \xrightarrow{p}\boldsymbol{\Omega}_0.
\end{aligned}
$$

$$
\begin{aligned}
(iii) \quad E[(d_{ii}^{-1} - \kappa_i^*/\tau_i^*)^2|\mathbf{Y}_T] &= \kappa_i^*/\tau_i^{*2} \\
&= \frac{\kappa_i + (T/2)}{\left[\tau_i + (T\mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i/2)\right]^2} \\
&= \frac{(\kappa_i/T) + (1/2)}{T\left[(\tau_i/T) + (\mathbf{a}_i'\hat{\mathbf{\Omega}}_{iT}^*\mathbf{a}_i/2)\right]^2} \\
&\xrightarrow{p} 0
\end{aligned}
$$

and

$$
\frac{\kappa_i^*}{\tau_i^*} \xrightarrow{p} \frac{(1/2)}{\mathbf{a}_i'\mathbf{\Omega}_0\mathbf{a}_i/2}.
$$

(iv) We first demonstrate that if $\mathbf{M}_i = \mathbf{M}$ for all i,

$$
\text{Prob}\{[\mathbf{A} \notin \mathbf{H}_\delta(\hat{\mathbf{\Omega}}_T^*)]|\mathbf{Y}_T\} \to 0 \qquad \forall \delta > 0. \tag{54}
$$

To see this, let $p_{ij}(\mathbf{A}, \mathbf{\Omega})$ denote the row i, column j element of $\mathbf{P}(\mathbf{A}, \mathbf{\Omega})$ for $\mathbf{P}(\mathbf{A}, \mathbf{\Omega})$ the lower-triangular Cholesky factor $\mathbf{P}(\mathbf{A}, \mathbf{\Omega})[\mathbf{P}(\mathbf{A}, \mathbf{\Omega})]' = \mathbf{A}\mathbf{\Omega}\mathbf{A}'$. Note that

$$
|\mathbf{A}\mathbf{\Omega}\mathbf{A}'| = p_{11}^2(\mathbf{A}, \mathbf{\Omega})p_{22}^2(\mathbf{A}, \mathbf{\Omega}) \cdots p_{nn}^2(\mathbf{A}, \mathbf{\Omega})
$$

$$
\mathbf{a}_i'\mathbf{\Omega}\mathbf{a}_i = p_{i1}^2(\mathbf{A}, \mathbf{\Omega}) + p_{i2}^2(\mathbf{A}, \mathbf{\Omega}) + \cdots + p_{ii}^2(\mathbf{A}, \mathbf{\Omega}).
$$

If $\mathbf{M}_i = \mathbf{M}$, then $\hat{\mathbf{\Omega}}_{iT}^* = \hat{\mathbf{\Omega}}_T^*$ and

$$
\begin{aligned}
p(\mathbf{A}|\mathbf{Y}_T) &= \frac{k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T^*\mathbf{A}')]^{T/2}}{\prod_{i=1}^n[(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_T^*\mathbf{a}_i]^{\kappa_i + T/2}} \\
&= \frac{k_T p(\mathbf{A})[p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)p_{22}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*) \cdots p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)]^{T/2}}{\prod_{i=1}^n[(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_T^*\mathbf{a}_i]^{\kappa_i}\left[(2\tau_i/T) + \sum_{j=1}^i p_{ij}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^{T/2}}.
\end{aligned}
$$

Note that for all $\mathbf{A} \notin H_\delta(\mathbf{\Omega}_T^*)$, $\exists\, j^* < i^*$ such that $\left[p_{i^*j^*}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^2 \geq \delta^*$ for $\delta^* = 2\delta/[n(n-1)]$.

Thus

$$\text{Prob}[\mathbf{A} \notin H_\delta(\hat{\mathbf{\Omega}}_T^*)|\mathbf{Y}_T] = \int_{\mathbf{A}\notin H_\delta(\hat{\mathbf{\Omega}}_T^*)} \frac{k_T p(\mathbf{A})[p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)p_{22}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\cdots p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)]^{T/2}}{\prod_{i=1}^n [(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_T^*\mathbf{a}_i]^{\kappa_i} \left[(2\tau_i/T) + \sum_{j=1}^i p_{ij}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^{T/2}} d\mathbf{A}$$

$$\leq \int_{\mathbf{A}\notin H_\delta(\hat{\mathbf{\Omega}}_T^*)} \left[\frac{k_T p(\mathbf{A})}{\prod_{i=1}^n [(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_T^*\mathbf{a}_i]^{\kappa_i}} \times \right.$$
$$\left. \frac{[p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)p_{22}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\cdots p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)]^{T/2}}{\prod_{i=1}^n [p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)]^{T/2}\cdots[\delta^* + p_{i^*i^*}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)]^{T/2}\cdots[p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)]^{T/2}}\right] d\mathbf{A}$$

$$= \int_{\mathbf{A}\notin H_\delta(\hat{\mathbf{\Omega}}_T^*)} \frac{k_T p(\mathbf{A})}{\prod_{i=1}^n [(2\tau_i/T) + \mathbf{a}_i'\hat{\mathbf{\Omega}}_T^*\mathbf{a}_i]^{\kappa_i}} \left[\frac{p_{i^*i^*}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)}{\delta^* + p_{i^*i^*}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)}\right]^{T/2} d\mathbf{A}$$

which goes to 0 as $T \to \infty$.

Note next that

$$\text{Prob}\{[\mathbf{A} \notin H_\delta(\mathbf{\Omega}_0)]|\mathbf{Y}_T\} = \text{Prob}\left[\sum_{i=2}^n \sum_{j=1}^{i-1} [p_{ij}(\mathbf{A}, \mathbf{\Omega}_0)]^2 > \delta\right].$$

But

$$\begin{aligned} [p_{ij}(\mathbf{A}, \mathbf{\Omega}_0)]^2 &= \left\{p_{ij}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*) + \left[p_{ij}(\mathbf{A}, \mathbf{\Omega}_0) - p_{ij}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]\right\}^2 \\ &\leq 2\left[p_{ij}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^2 + 2\left[p_{ij}(\mathbf{A}, \mathbf{\Omega}_0) - p_{ij}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^2. \end{aligned}$$

Hence

$$\text{Prob}\{[\mathbf{A} \notin H_\delta(\mathbf{\Omega}_0)]|\mathbf{Y}_T\} \leq \text{Prob}\{[(A_{1T} + A_{2T}) > \delta]|\mathbf{Y}_T\}$$

$$A_{1T} = 2\sum_{i=2}^n \sum_{j=1}^{i-1} \left[p_{ij}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^2$$

$$A_{2T} = 2\sum_{i=2}^n \sum_{j=1}^{i-1} \left[p_{ij}(\mathbf{A}, \mathbf{\Omega}_0) - p_{ij}(\mathbf{A}, \hat{\mathbf{\Omega}}_T^*)\right]^2.$$

Given any $\varepsilon > 0$ and $\delta > 0$, by virtue of (54) and result (ii) of Proposition 2, there exists a T_0 such that $\text{Prob}\{[A_{1T} > \delta/2]|\mathbf{Y}_T\} < \varepsilon/2$ and $\text{Prob}\{[A_{2T} > \delta/2]|\mathbf{Y}_T\} < \varepsilon/2$ for all $T \geq T_0$, establishing that $\text{Prob}\{[(A_{1T} + A_{2T}) > \delta]|\mathbf{Y}_T\} < \varepsilon$ as claimed.

(v) When $\kappa_i = \tau_i = 0$ and $\mathbf{M}_i = \mathbf{M}$, we have $\hat{\mathbf{\Omega}}^*_{iT} = \hat{\mathbf{\Omega}}^*_T$ and

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}^*_T\mathbf{A}')]^{T/2}}{\prod_{i=1}^n [\mathbf{a}_i'\hat{\mathbf{\Omega}}^*_T\mathbf{a}_i]^{T/2}}$$

which equals $k_T p(\mathbf{A})$ when evaluated at any $\mathbf{A}$ for which $\mathbf{A}\hat{\mathbf{\Omega}}^*_T\mathbf{A}'$ is diagonal.

D. Informative priors on lagged coefficients.

Doan, Litterman and Sims (1984) suggested that we should have greater confidence in our expectation that coefficients on higher lags are zero, represented by smaller diagonal elements for $\mathbf{M}_i$ associated with higher lags. For the rth element of $\mathbf{b}_i$, let $\ell(r)$ denote the lag order associated with that element,

$$\ell(r) = \begin{cases} 1 & \text{for } r = 1, ..., n \\ 2 & \text{for } r = n+1, ..., 2n \\ \vdots & \vdots \\ m & \text{for } r = n(r-1), ..., mn \end{cases},$$

and let $j(r)$ denote the explanatory variable associated with that coefficient,

$$j(r) = \begin{cases} 1 & \text{for } r = 1, n+1, ..., n(m-1)+1 \\ 2 & \text{for } r = 2, n+2, ..., n(m-1)+2 \\ \vdots & \vdots \\ n & \text{for } r = n-1, 2n-1, ..., mn \end{cases},$$

Let $\sqrt{s_{jj}}$ denote the estimated standard deviation of a univariate mth-order autoregression fit to variable j, and $M_{i,rr}$ the row r, column r element of $\mathbf{M}_i$. We then take[17]

$$\sqrt{M_{i,rr}} = \begin{cases} \frac{\lambda_0}{\sqrt{s_{ii}}[\ell(r)]^{\lambda_1}} & \text{if } i = j(r) \text{ and } r = 1, ..., k-1 \\ \frac{\lambda_0 \lambda_2}{\sqrt{s_{j(r),j(r)}}[\ell(r)]^{\lambda_1}} & \text{if } i \neq j(r) \text{ and } r = 1, ..., k-1 \\ \lambda_0 \lambda_3 & \text{for } r = k \end{cases} . \tag{55}$$

Here λ_0 summarizes the overall confidence in the prior (with smaller λ_0 corresponding to greater weight given to the random walk expectation), λ_1 governs how much more confident we are that higher coefficients are zero (with a value of $\lambda_1 = 0$ giving all lags equal weight), $\lambda_2 \leq 1$ is a parameter placing greater confidence in the restriction that coefficients other than the own lags $y_{i,t-\ell}$ are zero, and λ_3 is a separate parameter governing the tightness of the prior for the constant term, with all $\lambda_k \geq 0$.

E. Algorithm using the uniform Haar prior.

Here we describe the sign-restriction algorithm developed by Rubio-Ramírez, Waggoner, and Zha (2010) that was used to generate the histograms in Figure 5.

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 $\mathbf{K} = \mathbf{Q}' \cdot \mathbf{R}$ where $\mathbf{R}$ is an upper triangular matrix whose diagonal elements have been normalized to be positive and $\mathbf{Q}$ is an orthogonal matrix ($\mathbf{Q}\mathbf{Q}' = \mathbf{I}_n$). Let $\hat{\mathbf{P}}$ be the Cholesky factor of the reduced-form variance-covariance matrix $\hat{\mathbf{\Omega}}$ (so that $\hat{\mathbf{\Omega}} = \hat{\mathbf{P}}\hat{\mathbf{P}}'$) and generate a candidate impact matrix $\tilde{\mathbf{H}} = \hat{\mathbf{P}}\mathbf{Q}$. Instead of checking the sign restrictions directly for $\tilde{\mathbf{H}}$, normalize

[17] Doan (2013, p. UG-249) writes prior standard deviations as $\sqrt{s_{ii}}$ times the expressions in (55) to accommodate changes in the scale of variable i. However, in our parameterization this adjustment for scale is accommodated automatically by writing the variance of $\mathbf{b}_i$ as $d_{ii}\mathbf{M}_i$.

$\tilde{\mathbf{H}}$ by dividing each column by its first element as a way to account for both positive and negative shocks which increases the efficiency of the algorithm. Given that in sign-identified VARs the ordering of the variables does not determine which shock is contained in which column, each column needs to be checked for the sign pattern associated with one particular shock. If the normalized $\tilde{\mathbf{H}}$ satisfies all the sign restrictions jointly, keep the draw; otherwise discard it.

F. A new-Keynesian interpretation of the VAR in Section 5.

Benati and Surico (2009) used the following New Keynesian model:

$$y_t = \eta y_{t+1|t} + (1-\eta)y_{t-1} - \sigma^{-1}(r_t - \pi_{t+1|t}) + \varepsilon_{yt} \tag{56}$$

$$\varepsilon_{yt} = \rho_y \varepsilon_{y,t-1} + \tilde{\varepsilon}_{yt} \tag{57}$$

$$\pi_t = \xi \pi_{t+1|t} + \alpha^s \pi_{t-1} + \kappa^s y_t + \tilde{\varepsilon}_{\pi t} \tag{58}$$

$$r_t = \rho r_{t-1} + \psi^y y_t + \psi^\pi \pi_t + \varepsilon_{rt} \tag{59}$$

$$\varepsilon_{rt} = \rho_r \varepsilon_{r,t-1} + \tilde{\varepsilon}_{rt}$$

where $(\tilde{\varepsilon}_{yt}, \tilde{\varepsilon}_{\pi t}, \tilde{\varepsilon}_{rt})'$ is vector white noise and $x_{t+1|t}$ denotes the rational expectation of x_{t+1} formed on the basis of information available at time t. If we operate on (59) by $(1-\rho_r L)$ for L the lag operator, we obtain an equation of the form of (46) with $u_t^m = \tilde{\varepsilon}_{rt}$, and in the special case of a backward-looking Phillips Curve ($\xi = 0$), equation (58) is immediately in the form of (44) with $\kappa^s = 1/\alpha^s$. Benati and Surico show that the rational-expectations solution to their system takes the form

$$y_{t+1|t} = d_y y_t + d_\pi \pi_t + d_r r_t + \mathbf{d}'\mathbf{y}_{t-1} \tag{60}$$

$$\pi_{t+1|t} = g_y y_t + g_\pi \pi_t + g_r r_t + \mathbf{g}'\mathbf{y}_{t-1}. \tag{61}$$

Substituting (60) and (61) into (56) and rearranging gives an expression of the form of (45).

References

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2013). "Algorithm for Inference with Sign and Zero Restrictions," working paper, Duke University.

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

Caldara, Dario and Christophe Kamps (2012). "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers," working paper, Federal Reserve Board.

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

Canova, Fabio, and Matthias Paustian (2011). "Business Cycle Measurement with Some Theory," *Journal of Monetary Economics* 58: 345-361.

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

Doan, Thomas (2013). *RATS User's Guide, Version 8.2,* www.estima.com.

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

Elliott, Graham, Ulrich K. Müller, and Mark W. Watson (2012). "Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis," working paper, Princeton University.

Faust, Jon (1998). "The Robustness of Identified VAR Conclusions about Money," *Carnegie-Rochester Series on Public Policy* 49: 207-244.

Ferguson, Thomas S. (1967), *Mathematical Statistics: A Decision Theoretic Approach,* New York: Academic Press.

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

Giacomini, Raffaella, and Toru Kitagawa (2013). "Robust Bayes Inference for Partially Identified VARs," working paper, University College London.

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

Gubner, John A. (2006). *Probability and Random Processes for Electrical and Computer Engineers.* Cambridge: Cambridge University Press.

Haar, A. (1933). "Der Massbegriff in der Theorie der kontinuierlichen Gruppen", *Annals of Mathematics* second series 34(1): 147–169.

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

Juvenal, Luciana, and Ivan Petrella (2011). "Speculation in the Oil Market," forthcoming, *Journal of Applied Econometrics.*

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

Kilian, Lutz, and Daniel P. Murphy (2012). "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," *Journal of the European Economic Association* 10(5): 1166-1188.

Leamer, Edward E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: John Wiley & Sons, Inc.

Litterman, Robert B. (1986). "Forecasting With Bayesian Vector Autoregressions– Five Years of Experience," *Journal of Business & Economic Statistics* 4: 25-38.

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

Moon, Hyungsik Roger, and Frank Schorfheide (2012). "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica* 80: 755-782.

Moon, Hyungsik Roger, Frank Schorfheide, and Eleonora Granziera (2013). "Inference for VARs Identified with Sign Restrictions", working paper, University of Pennsylvania.

Müller, Ulrich K. and Andriy Norets (2012). "Credibility of Confidence Sets in Nonstandard Econometric Problems," working paper, Princeton University.

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

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.

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.

Shapiro, Matthew D., and Mark W. Watson (1988). "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual*, Volume 3, pp. 111-156, edited by Stanley Fischer. Cambridge, MA: M.I.T. Press.

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

Smets, Frank and Raf Wouters (2003). "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association* 1: 1123–1175.

Stewart, G.W. (1980). "The Efficient Generation of Random Orthogonal Matrices with an Application to Conditional Estimators," *SIAM Journal on Numerical Analysis* 17(3): 403-409.

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 (c)	Sign restriction
α^s	Effect of π on supply	2	$\alpha^s \geq 0$
β^d	Effect of π on demand	0.75	none
γ^d	Effect of r on demand	-1	none
ψ^y	Fed response to y	0.5	$\psi^y \geq 0$
ψ^π	Fed response to π	1.5	$\psi^\pi \geq 0$

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 shock*	
	Prior	Posterior	Prior	Posterior	Prior	Posterior
Variable						
			s = 0			
$\boldsymbol{y}$	0.814	1.000	0.974	1.000	0.024	0.005
$\boldsymbol{\pi}$	0.019	0.000	0.974	1.000	0.024	0.005
$\boldsymbol{r}$	0.032	0.148	0.974	1.000	0.938	1.000
			s = 1			
$\boldsymbol{y}$	0.755	1.000	0.930	1.000	0.046	0.015
$\boldsymbol{\pi}$	0.051	0.000	0.858	1.000	0.089	0.009
$\boldsymbol{r}$	0.087	0.335	0.928	1.000	0.852	0.995
			s = 2			
$\boldsymbol{y}$	0.680	1.000	0.842	1.000	0.102	0.056
$\boldsymbol{\pi}$	0.086	0.000	0.750	1.000	0.208	0.016
$\boldsymbol{r}$	0.141	0.483	0.882	1.000	0.760	0.990

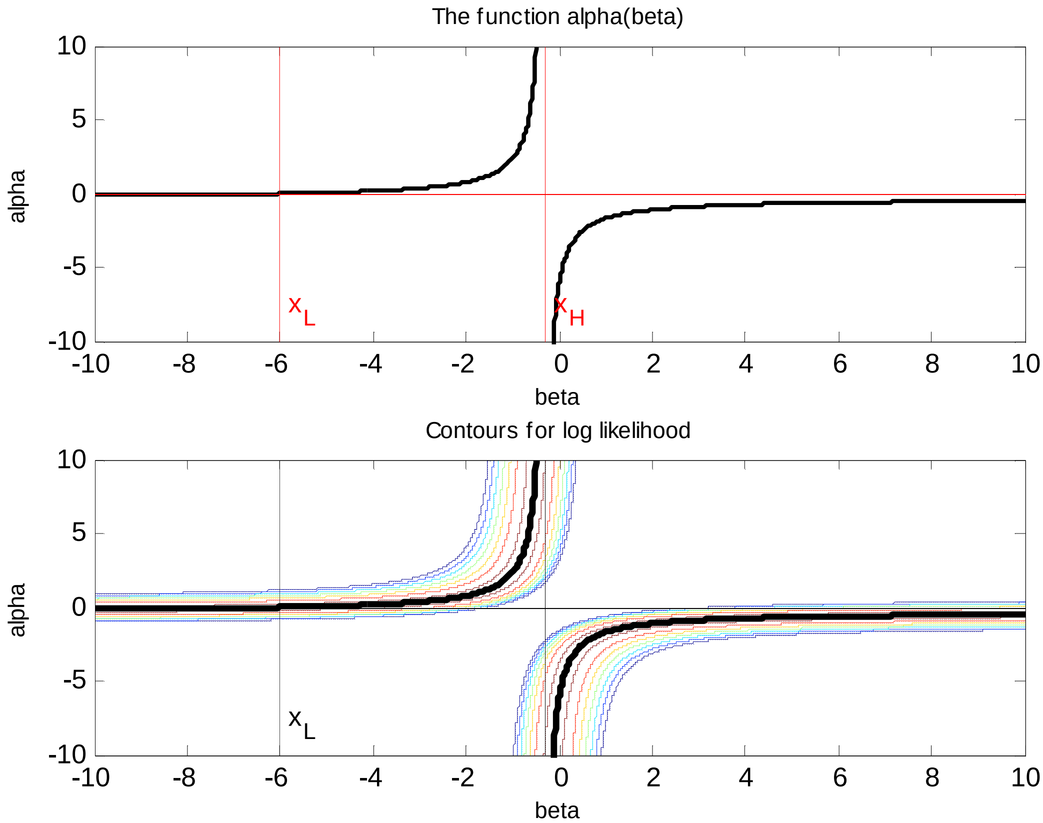

Figure 1. Maximum likelihood estimates and likelihood contours for α and β. Distance between contour lines is 3, and unshaded regions are exceedingly unlikely given the data.

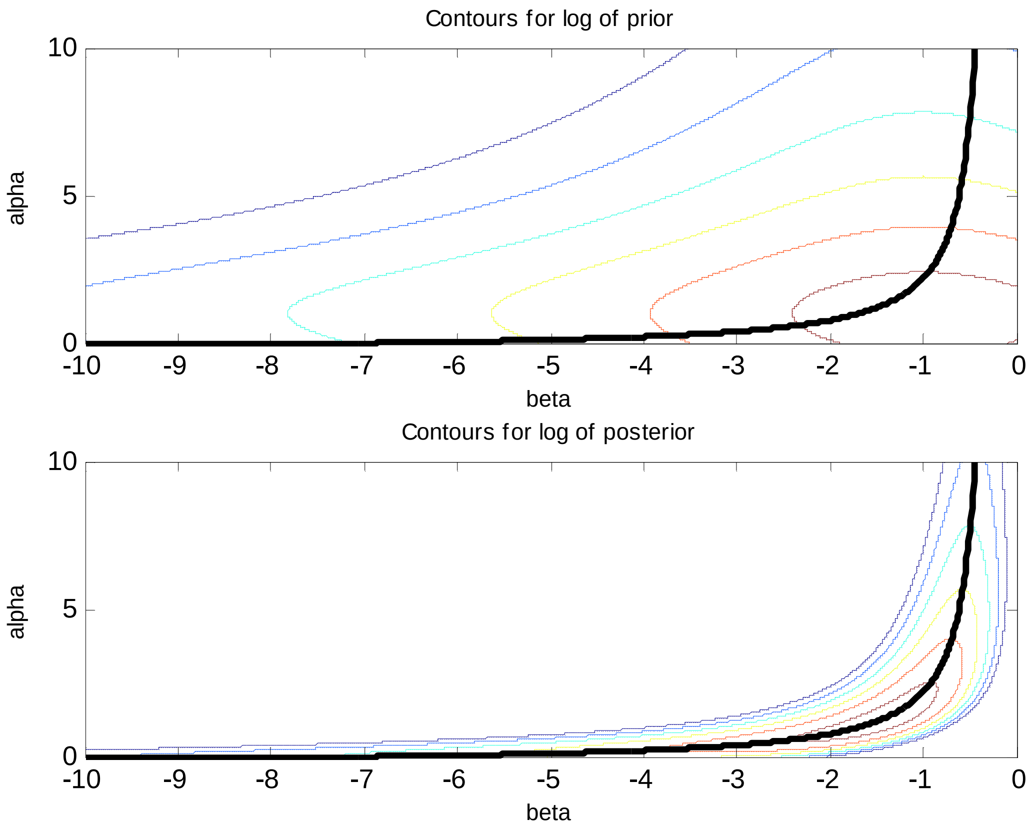

Figure 2. Maximum likelihood estimates and contours of prior and posterior distribution. Distance between contour lines is 3, and unshaded regions are exceedingly unlikely given prior beliefs and the data.

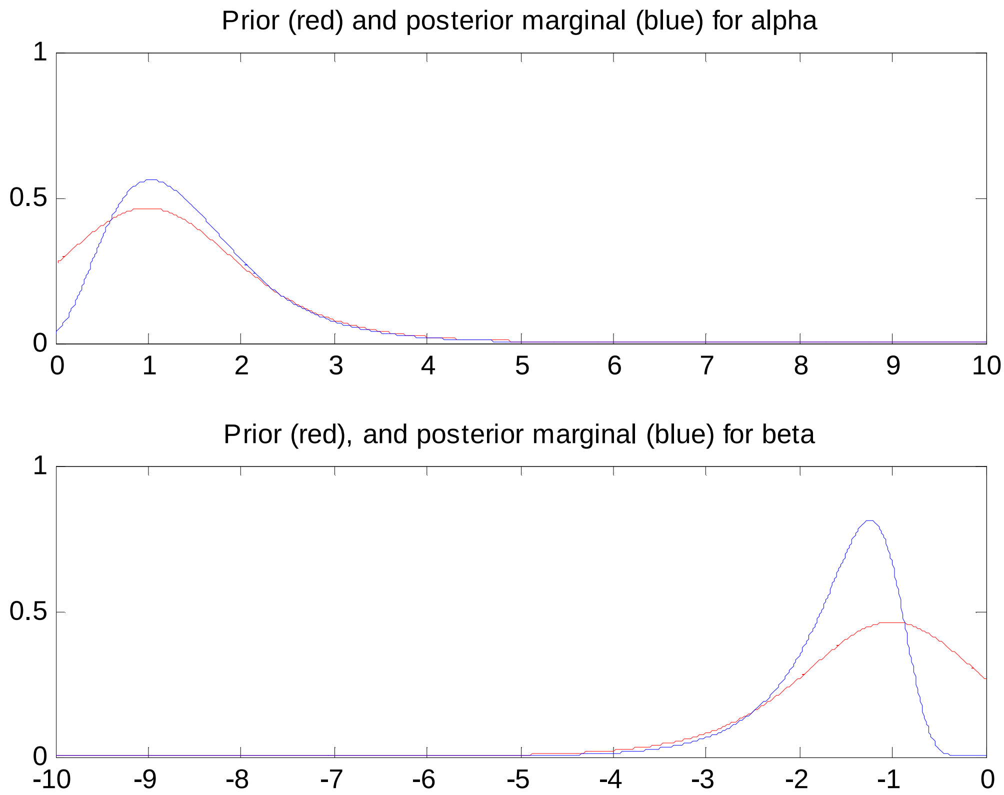

Figure 3. Prior and posterior distributions for α and β.

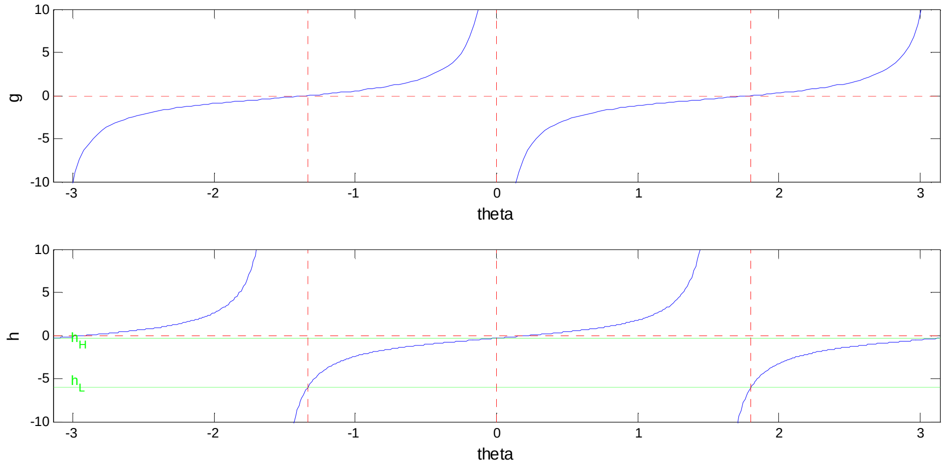

Figure 4. Graphs of equations (32) and (33).

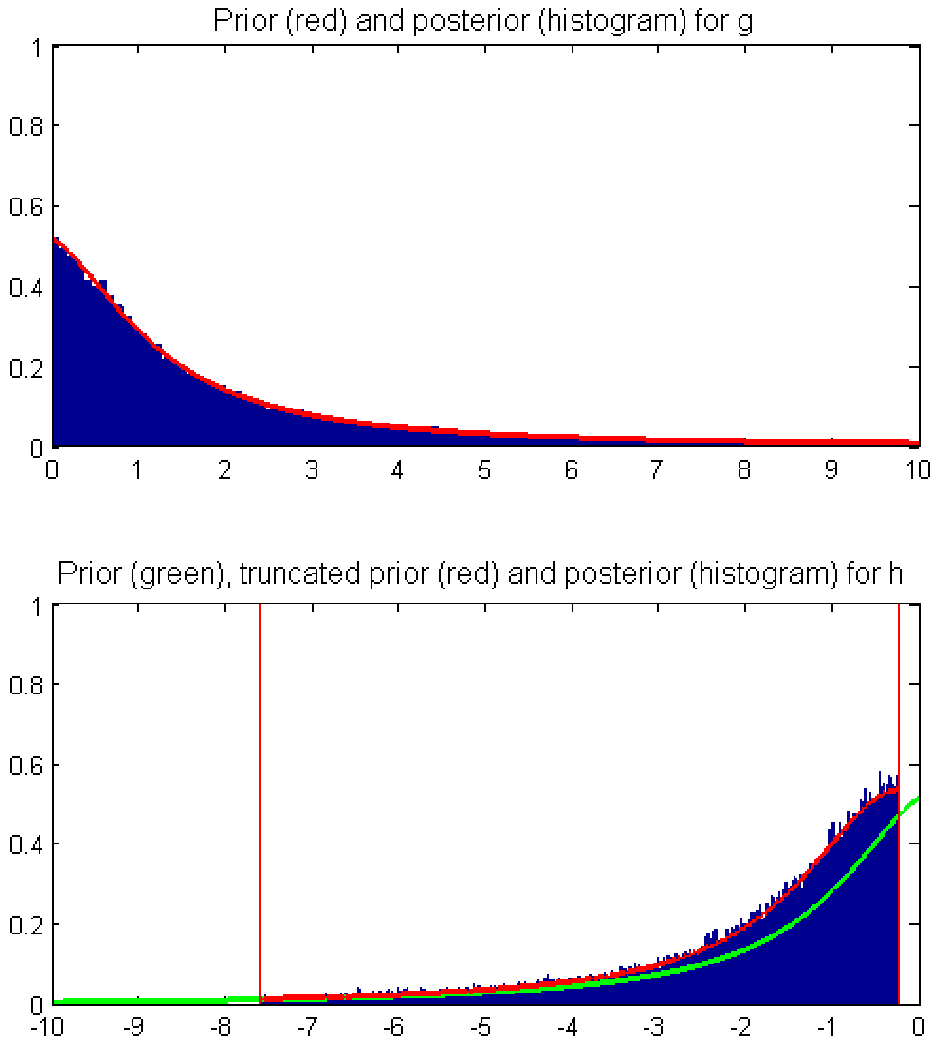

Figure 5. Blue bars represented estimated histogram for the date zero impact of a demand shock (top panel) and supply shock (bottom panel) as inferred from applying the conventional sign-restriction methodology to postwar U.S. data. Red curve in top panel represents the Cauchy prior truncated by $g > 0$, while red curve in bottom panel represents the Cauchy prior truncated by $x_L < h < x_H$.

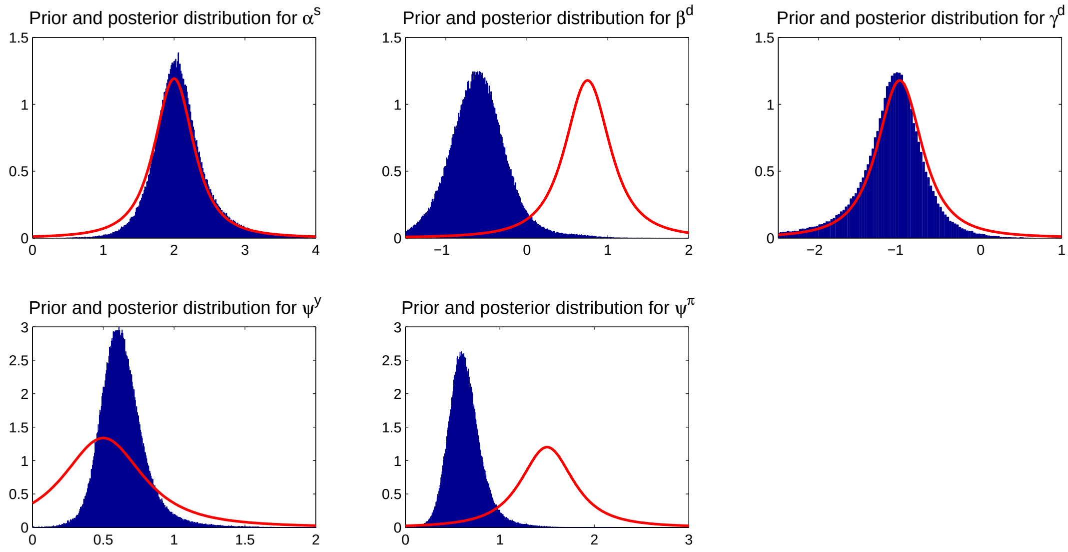

Figure 6. Prior distributions (in red) and posterior distributions (blue histogram) for contemporaneous parameters for 3-variable macro model.

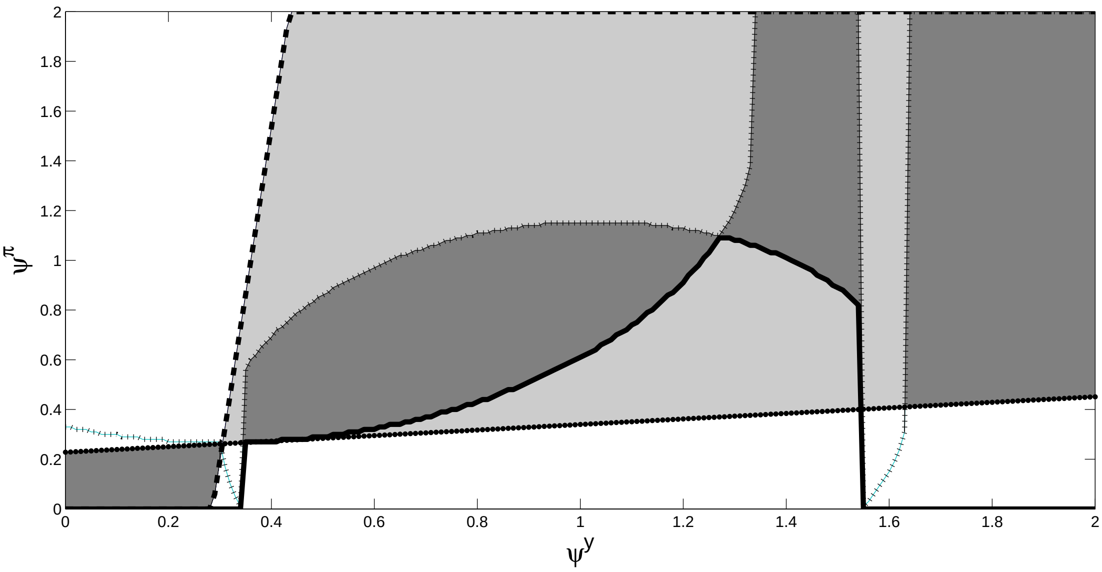

Figure 7. Implications of different values for ψ^y and ψ^π. Dashed line: values for which MLE of α^s is zero. Beaded line: values for which MLE of α^s is infinity. Dark shaded region: $\alpha^s > 0$ and $\beta^d < 0$. Light shaded region: $\alpha^s > 0$ and $\beta^d > 0$.

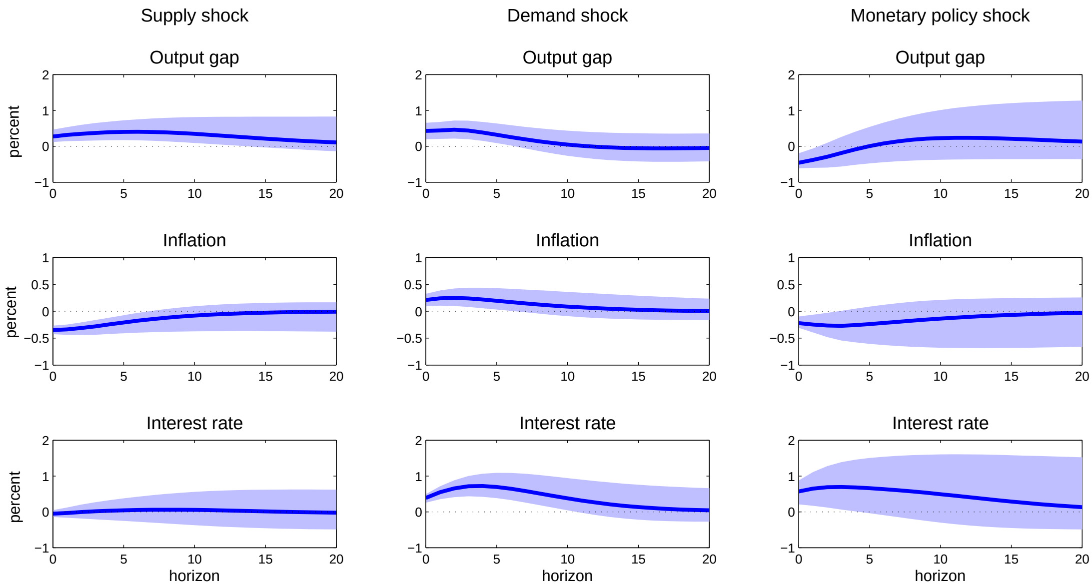

Figure 8. Impulse-response functions for 3-variable VAR with informative priors for lagged coefficients. Solid lines: posterior median. Shaded regions: 95% posterior credibility set.

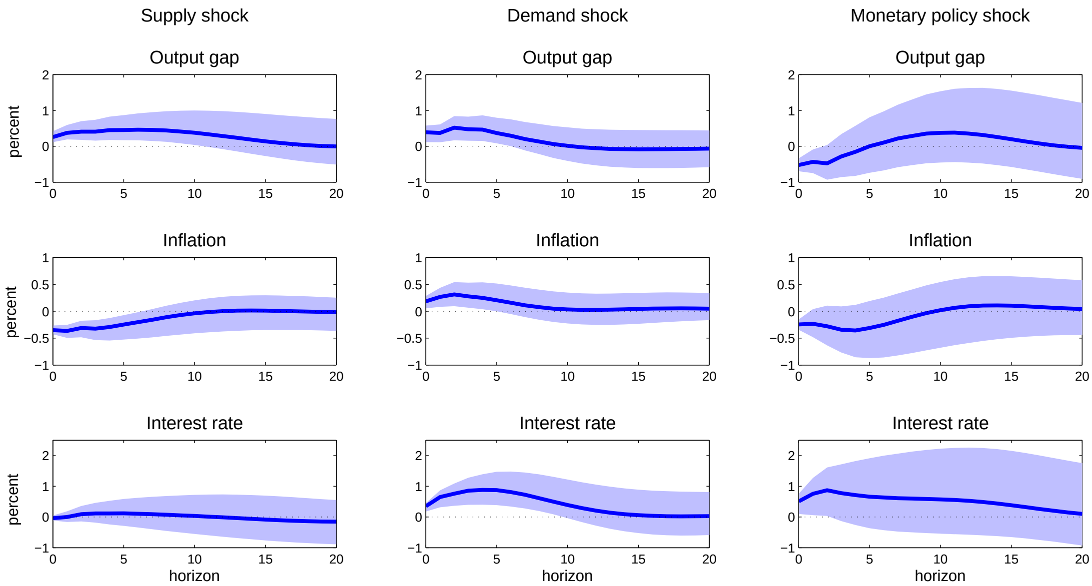

Figure 9. Impulse-response functions for 3-variable VAR with noninformative priors for lagged coefficients. Solid lines: posterior median. Shaded regions: 95% posterior credibility set.