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: December 3, 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 Luc Bauwens, Fabio Canova, Gary Chamberlain, Drew Creal, Lutz Kilian, Adrian Pagan, Elie Tamer, Harald Uhlig, and anonymous referees for helpful comments.

ABSTRACT

This paper makes the following original contributions to the literature. (1) We develop a simpler analytical characterization and numerical algorithm for Bayesian inference in structural vector autoregressions that can be used for models that are overidentified, just-identified, or underidentified. (2) We analyze the asymptotic properties of Bayesian inference and show that in the underidentified case, 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. For example, in a bivariate VAR for supply and demand identified solely by sign restrictions, if the population correlation between the VAR residuals is positive, then even if one has available an infinite sample of data, any inference about the demand elasticity is coming exclusively from the prior distribution. (3) We provide analytical characterizations of the informative prior distributions for impulse-response functions that are implicit in the traditional sign-restriction approach to VARs, and note, as a special case of result (2), that the influence of these priors does not vanish asymptotically. (4) We illustrate how Bayesian inference with informative priors can be both a strict generalization and an unambiguous improvement over frequentist inference in just-identified models. (5) We propose that researchers need to explicitly acknowledge and defend the role of prior beliefs in influencing structural conclusions and illustrate how this could be done using a simple model of the U.S. labor market.

1 Introduction.

In pioneering papers, Blanchard and Diamond (1990), Faust (1998), Davis and Haltiwanger (1999), 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.

But 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.

There is a huge literature that considers econometric inference under set identification using a frequentist approach; see for example the reviews in Manski (2003) and Tamer (2010). However, to our knowledge Moon, Schorfheide and Granziera (2013) is the only effort to apply these methods to sign-restricted VARs, where the number of parameters can be very large and the topology of the identified set quite complex. Instead, the hundreds of researchers who have estimated sign-restricted VARs have virtually all used numerical methods that are essentially Bayesian in character, though often without acknowledging that the methods represent an application of Bayesian principles.

If the data are uninformative for distinguishing between elements within a set, for some questions of interest the Bayesian posterior inference will continue to be influenced by prior beliefs even if the sample size is infinite. This point has been well understood in the literature on Bayesian inference in set-identified models; see for example Poirier (1998), Gustafson (2009), and Moon and Schorfheide (2012). However, the implications of this fact for the procedures popularly used for set-identified VARs have not been previously documented.

The popular numerical algorithms currently in use for sign-identified VARs only work for a very particular prior distribution, namely the uniform Haar prior. By contrast, in this paper we provide analytical results and numerical algorithms for Bayesian inference for a quite general class of prior distributions for structural VARs. Our expressions simplify the methods proposed by Sims and Zha (1998) as well as generalize them to the case in which the structural assumptions may not be sufficient to achieve full identification. Our formulation also allows us to characterize the asymptotic properties of Bayesian inference in structural VARs that may not be fully identified.

We demonstrate that as the sample size goes to infinity, the analyst could know with certainty that contemporaneous structural coefficients fall within a set $S(\mathbf{\Omega})$ that orthogonalizes the true variance-covariance matrix, but within this set, the height of the posterior distribution is simply a constant times the height of the prior distribution at that point. In the case of a bivariate model of supply and demand in which sign restrictions are the sole identifying assumption, if the reduced-form residuals have positive correlation, then $S(\mathbf{\Omega})$ allows any value for the elasticity of demand but restricts the elasticity of supply to fall within a particular interval. With negatively correlated errors, the elasticity of supply could be any positive number while the elasticity of demand is restricted to fall in a particular interval.

We also explore the implications of the Haar prior. We demonstrate that although this is commonly regarded as uninformative, in fact it implies nonuniform distributions for key objects of interest. It implies that the impact of a one-standard-deviation structural shock is regarded before seeing the data as coming from a distribution with more mass at the center when the number of variables n in the VAR is greater than 3 but more mass at the extremes when $n = 2$. We also show that the Haar distribution implies Cauchy priors for structural parameters such as elasticities. We demonstrate that users of these methods can in some cases end up performing hundreds of thousands of calculations, ostensibly analyzing the data, but in fact are doing nothing more than generating draws from a prior distribution that they never even acknowledged assuming.

We recommend instead that any prior beliefs should be acknowledged and defended openly and their role in influencing posterior conclusions clearly identified. We claim a number of advantages of this approach over existing methods for both underidentified and just-identified VARs. First, although sign restrictions only achieve set-identification, most users of these methods are tempted to summarize their results using point estimates, an approach that is deeply problematic from a frequentist perspective (see for example Fry and Pagan, 2011). By contrast, if priors accurately reflect our uncertainty about parameters and the underlying structure, then given a loss function there is an unambiguously optimal posterior estimate to report for any object of interest, and the Bayesian posterior distribution accurately represents our combined uncertainty resulting from having a limited set of observed data as well as possible doubts about the true structure itself. Second, researchers like Kilian and Murphy (2012) and Caldara and Kamps (2012) have argued persuasively for the benefits of using additional prior information about parameters as a supplement to sign restrictions in VARs. Our approach provides a flexible and formal apparatus for doing exactly this in quite general settings. Third, our proposed Bayesian methods can be viewed as a strict generalization of the conventional approach to identification, with multiple advantages, including better statistical treatment of joint uncertainty about contemporaneous and lagged coefficients as well as the opportunity to visualize, as in Leamer (1981), the consequences for the posterior inference of relaxing the role of the prior.[1]

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 provides an analytical characterization of the priors that are implicit in the popular approach to sign-identified VARs. Section 4 discusses the use of additional information about impacts at longer horizons, noting the need to formulate these in terms of joint prior beliefs about contemporaneous and lagged structural coefficients. Section 5 illustrates our recommended methods using a simple model of the U.S. labor market. Section 6 briefly concludes.

[1] 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.

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

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

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

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

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

Note that maximum likelihood estimates of the reduced-form parameters are given by

$$\hat{\mathbf{\Phi}}_T = \left(\textstyle\sum_{t=1}^T \mathbf{y}_t\mathbf{x}'_{t-1}\right)\left(\textstyle\sum_{t=1}^T \mathbf{x}_{t-1}\mathbf{x}'_{t-1}\right)^{-1} \tag{6}$$

$$\hat{\mathbf{\Omega}}_T = T^{-1}\textstyle\sum_{t=1}^T \hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}'_t \tag{7}$$

for $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{\Phi}}_T\mathbf{x}_{t-1}$.

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.[2]

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 prior information about $\mathbf{D}$ and $\mathbf{B}$ we employ natural conjugate distributions which facilitate analytical characterization of results as well as 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,[3]

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

[2] Our derivations draw on insights from Sims and Zha (1998). The main difference is that they parameterize the contemporaneous relations in terms of a single matrix, whereas we use two matrices $\mathbf{A}$ and $\mathbf{D}$, and take advantage of the fact that the posterior distribution of $\mathbf{D}$ is known analytically. As a consequence the core expression in our result (equation (21)) is simpler than their equation (10). Among other benefits, we show that the asymptotic properties of (21) can be obtained analytically.

[3] We will follow the notational convention of using $p(.)$ to denote any 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 $p_{\mathbf{D}}(\mathbf{D})$.

$$p(d_{ii}^{-1}|\mathbf{A}) = \begin{cases} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}(d_{ii}^{-1})^{\kappa_i-1}\exp(-\tau_i d_{ii}^{-1}) & \text{for } d_{ii}^{-1} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where d_{ii} denotes the (i,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. Our general results below allow both κ_i and τ_i to be arbitrary functions of $\mathbf{A}$.

Normal priors are used for the lagged structural coefficients $\mathbf{B}$, with results particularly simple if coefficients are taken to be independent $N(\mathbf{m}_i, d_{ii}\mathbf{M}_i)$ across equations:

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

$$p(\mathbf{b}_i|\mathbf{D},\mathbf{A}) = \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}}\exp[-(1/2)(\mathbf{b}_i - \mathbf{m}_i)'(d_{ii}\mathbf{M}_i)^{-1}(\mathbf{b}_i - \mathbf{m}_i)]. \tag{10}$$

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. We allow $\mathbf{m}_i$ and $\mathbf{M}_i$ to be functions of $\mathbf{A}$ but not of $\mathbf{D}$.

The overall prior is then

$$p(\mathbf{A},\mathbf{D},\mathbf{B}) = p(\mathbf{A})\textstyle\prod_{i=1}^n[p(d_{ii}|\mathbf{A})p(\mathbf{b}_i|\mathbf{D},\mathbf{A})]. \tag{11}$$

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

$$\begin{aligned} p(\mathbf{Y}_T|\mathbf{A},\mathbf{D},\mathbf{B}) &= (2\pi)^{-Tn/2}|\det(\mathbf{A})|^T|\mathbf{D}|^{-T/2} \times \\ &\quad \exp\left[-(1/2)\textstyle\sum_{t=1}^T(\mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1})'\mathbf{D}^{-1}(\mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1})\right] \end{aligned} \tag{12}$$

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

Components of the Bayesian posterior distributions can be conveniently characterized by regressions on augmented data sets defined by

$$\underset{[(T+k)\times 1]}{\tilde{\mathbf{Y}}_i} = \begin{bmatrix} \mathbf{y}_1'\mathbf{a}_i \\ \vdots \\ \mathbf{y}_T'\mathbf{a}_i \\ \mathbf{P}_i'\mathbf{m}_i \end{bmatrix} \tag{13}$$

$$\underset{[(T+k)\times k]}{\tilde{\mathbf{X}}_i} = \begin{bmatrix} \mathbf{x}_0' \\ \vdots \\ \mathbf{x}_{T-1}' \\ \mathbf{P}_i' \end{bmatrix} \tag{14}$$

for $\mathbf{P}_i$ the Cholesky factor of $\mathbf{M}_i^{-1} = \mathbf{P}_i\mathbf{P}_i'$. 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 (12) and priors are given by (8)-(11), then for $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{X}}_i$ defined by (13) and (14), 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

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

$$\mathbf{m}_i^* = \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i\right) \tag{15}$$

$$\mathbf{M}_i^* = \left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1} \tag{16}$$

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

$$\kappa_i^* = \kappa_i + (T/2) \tag{17}$$

$$\tau_i^* = \tau_i + (\zeta_i^*/2) \tag{18}$$

$$\zeta_i^* = \left(\tilde{\mathbf{Y}}_i'\tilde{\mathbf{Y}}_i\right) - \left(\tilde{\mathbf{Y}}_i'\tilde{\mathbf{X}}_i\right)\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i\right) \tag{19}$$

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')]^{T/2}}{\prod_{i=1}^n[(2\tau_i^*/T)]^{\kappa_i^*}}\prod_{i=1}^n\left\{\frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}}\frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}\Gamma(\kappa_i^*)\right\} \tag{20}$$

for $\hat{\mathbf{\Omega}}_T$ *given by* (7) *and* k_T *the constant for which (20) integrates to unity.*

Note that if the prior parameters $\mathbf{M}_i$, κ_i and τ_i do not depend on $\mathbf{A}$, the last term in (20) can be subsumed into the constant term k_T, in which case (20) simplifies to[4]

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')]^{T/2}}{\prod_{i=1}^n[(2\tau_i^*/T)]^{\kappa_i^*}}. \tag{21}$$

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

[4] Note also for the special case of noninformative priors, as $\kappa_i, \tau_i \to 0$, $\tau_i^{\kappa_i} \to 1$.

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,

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

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 $\mathbf{M}_i^{-1} = \mathbf{0}$) we have that $\zeta_i^* = T\mathbf{a}_i'\hat{\mathbf{\Omega}}_T\mathbf{a}_i$. 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 ζ_i^*/T still converges to $\mathbf{a}_i'\mathbf{\Omega}_0\mathbf{a}_i$ for $\mathbf{\Omega}_0 = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$ the true variance matrix, and the Bayesian posterior distribution for d_{ii}^{-1} conditional on $\mathbf{A}$ collapses to a point mass at $1/(\mathbf{a}_i'\mathbf{\Omega}_0\mathbf{a}_i)$. 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 (20) simplifies to

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A})|\det(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')|^{T/2}}{\left\{\det\left[\mathrm{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')\right]\right\}^{T/2}} \tag{22}$$

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

$$S(\mathbf{\Omega}) = \{\mathbf{A} : \mathbf{A}\mathbf{\Omega}\mathbf{A}' = \mathrm{diag}(\mathbf{A}\mathbf{\Omega}\mathbf{A}')\}, \tag{23}$$

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 $\mathbf{A\Omega A}'$,

$$q(\mathbf{A},\mathbf{\Omega}) = \textstyle\sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ij}^2(\mathbf{A},\mathbf{\Omega}) \qquad \mathbf{P}(\mathbf{A},\mathbf{\Omega})[\mathbf{P}(\mathbf{A},\mathbf{\Omega})]' = \mathbf{A\Omega A}', \tag{24}$$

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

$$H_\delta(\mathbf{\Omega}) = \{\mathbf{A} : q(\mathbf{A},\mathbf{\Omega}) \leq \delta\}. \tag{25}$$

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*

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

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

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*

$$\begin{aligned} p(\mathbf{A}, d_{11}^{-1}, ..., d_{nn}^{-1}, \mathbf{B}|\mathbf{Y}_T) \;&=\; k_T p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_T\mathbf{A}')]^{T/2} \times \\ &\quad \textstyle\prod_{i=1}^{n} \dfrac{\gamma(d_{ii}^{-1};\kappa_i^*,\tau_i^*)\phi(\mathbf{b}_i;\mathbf{m}_i^*,d_{ii}\mathbf{M}_i^*)}{[(2\tau_i^*/T)]^{\kappa_i^*}} \end{aligned} \tag{26}$$

where $\mathbf{a}_i'$ *and* $\mathbf{b}_i'$ *denote the* i*th rows of* $\mathbf{A}$ *and* $\mathbf{B}$, *respectively, and* $\kappa_i^*, \tau_i^*, \mathbf{m}_i^*, \mathbf{M}_i^*, \zeta_i^*$ *and* $\hat{\mathbf{\Omega}}_T$ *are the functions of* $\mathbf{Y}_T$ *defined in Proposition 1 where* $\mathbf{M}_i$ *is any invertible* $(k \times k)$

matrix, $\mathbf{m}_i$ *any finite* $(k \times 1)$ *vector,* τ_i *and* κ_i *any finite nonnegative constants, and* k_T *an integrating constant that depends only on* $\hat{\mathbf{\Omega}}_T, \tau_i, \kappa_i$ $(i = 1, ..., n)$ *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 (25). Then as the sample size* T *goes to infinity, the random variables characterized by (26) have the following properties:*

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

(*ii*) $(\zeta_i^*/T)|\mathbf{A}, \mathbf{Y}_T \xrightarrow{p} \mathbf{a}_i'\mathbf{\Omega}_0\mathbf{a}_i$;

(*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 $\mathbf{M}_i = \mathbf{0}$ 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 (26) 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 $\mathbf{A}\boldsymbol{\Omega}_0\mathbf{A}'$ is diagonal. Provided that $p(\mathbf{A})$ is nonzero in all neighborhoods 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.

We close this section with comments on some alternative ways we could parameterize the model. One possibility would be to work directly with the reduced form in (2),

$$\mathbf{y}_t = \boldsymbol{\Phi}\mathbf{x}_{t-1} + \tilde{\mathbf{H}}\mathbf{u}_t, \tag{27}$$

and specify priors over $\tilde{\mathbf{H}}$, the matrix of date-zero impacts of structural shocks,

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

Recall from (4) that $\tilde{\mathbf{H}} = \mathbf{A}^{-1}$. Since Proposition 1 established results for an arbitrary set of prior beliefs about $\mathbf{A}$, those expressions also immediately characterize the posterior inference that would emerge from any set of prior beliefs about $\tilde{\mathbf{H}}$. Alternatively we could consider forming priors directly over the impacts of one-standard-deviation shocks: $\mathbf{H} = \mathbf{A}^{-1}\mathbf{D}^{1/2}$.

One potential advantage that might be claimed for the latter approaches is that if we are interested only in the effects of the jth structural shock, we only need information about the jth column of $\tilde{\mathbf{H}}$ or $\mathbf{H}$ rather than the entire matrix $\mathbf{A}$ as in our proposed treatment. However, in our view it would be unusual to know nothing about many of the specific structural relations in $\mathbf{A}$ and yet have confidence about how those different equations interact to produce equilibrium outcomes summarized by certain elements of $\mathbf{A}^{-1}$. The vast majority of useful prior economic information (for example, all of the hundreds of studies on which we draw in our empirical application in Section 5) have been formulated in terms of things we know about $\mathbf{A}$ rather than $\mathbf{A}^{-1}$. Simulations by Canova and Paustian (2011) also document practical limitations to trying to draw inference using restrictions on only one column of $\mathbf{H}$ even if those restrictions are known to be correct. For this reason we believe that the parameterization used in Propositions 1 and 2 is preferred.

Yet another approach would be to calculate the prior distribution for the reduced-form parameters $\{\mathbf{\Phi}, \mathbf{\Omega}\}$ that is implied by our prior $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$ and then rewrite the latter as $p(\mathbf{A}|\mathbf{\Phi}, \mathbf{\Omega})p(\mathbf{\Phi}, \mathbf{\Omega})$. For such a parameterization we know as in Drèze (1974) that the posterior $p(\mathbf{\Phi}, \mathbf{\Omega}|\mathbf{Y}_T)$ will become perfectly informative about the reduced-form parameters while $p(\mathbf{A}|\mathbf{\Phi}, \mathbf{\Omega}, \mathbf{Y}_T) = p(\mathbf{A}|\mathbf{\Phi}, \mathbf{\Omega})$. Again the practical challenge is how to characterize the implications of previous research and prior knowledge in terms of a distribution $p(\mathbf{A}|\mathbf{\Phi}, \mathbf{\Omega})$. For example, we are aware of no empirical studies that have actually implemented Drèze's suggestion in the half-century since it was first proposed for purposes of analyzing systems of simultaneous equations. While one could use numerical methods to characterize the distributions $p(\mathbf{\Phi}, \mathbf{\Omega})$ and $p(\mathbf{A}|\mathbf{\Phi}, \mathbf{\Omega})$ that are implicit in our suggested priors, our impression is that calculating these would be more computationally taxing than the approach spelled out in Appendix B, most of whose key steps are known analytically.

3 Priors implicit in the traditional approach to sign-identified VARs.

In this section we characterize the prior distributions that are implicit in the approach to sign-restricted VARs developed by Rubio-Ramírez, Waggoner, and Zha (2010). Their algorithm begins by generating an $(n \times n)$ matrix $\mathbf{X} = [x_{ij}]$ of independent $N(0,1)$ variates and then calculating the decomposition $\mathbf{X} = \mathbf{QR}$ where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is upper triangular. From this a candidate draw for the contemporaneous response to the vector of structural shocks is calculated as

$$\mathbf{H} = \mathbf{PQ} \tag{28}$$

for $\mathbf{P}$ the Cholesky factor of the variance matrix of the reduced-form innovations:

$$\mathbf{\Omega} = \mathbf{PP}'. \tag{29}$$

This algorithm can be viewed as generating draws from a Bayesian prior distribution for $\mathbf{H}$ conditional on $\mathbf{\Omega}$, the properties of which we now document. Note that the first column of $\mathbf{Q}$ is simply the first column of $\mathbf{X}$ normalized to have unit length:

$$\begin{bmatrix} q_{11} \\ \vdots \\ q_{n1} \end{bmatrix} = \begin{bmatrix} x_{11}/\sqrt{x_{11}^2 + \cdots + x_{n1}^2} \\ \vdots \\ x_{n1}/\sqrt{x_{11}^2 + \cdots + x_{n1}^2} \end{bmatrix}. \tag{30}$$

In the special case when $n = 2$, q_{11} has the interpretation as the cosine of θ, the angle defined by the point (x_{11}, x_{21}). In this case, the fact that $\mathbf{Q}$ is an orthogonal matrix will result in half the draws being the rotation matrix associated with θ and the other half the reflection matrix:

$$\mathbf{Q} = \begin{cases} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} & \text{with probability } 1/2 \\ \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} & \text{with probability } 1/2 \end{cases}. \tag{31}$$

Independence of x_{11} and x_{21} ensures that θ will have a uniform distribution over $[-\pi, \pi]$. It is in this sense that the prior implicit in the procedure (28) is thought to be uninformative.

However, a prior that is uninformative about some features of the parameter space can turn out to be informative about others. It is not hard to show that each element of (30) has a marginal density given by[5]

$$p(q_{i1}) = \begin{cases} \frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)}(1-q_{i1}^2)^{(n-3)/2} & \text{if } q_{i1} \in [-1,1] \\ 0 & \text{otherwise} \end{cases}. \tag{32}$$

Note next that the $(1,1)$ element of (28) therefore implies a prior distribution for the effect of a one-standard-deviation increase in structural shock number 1 on variable number 1 that is characterized by the random variable

$$h_{11} = p_{11}q_{11} = \sqrt{\omega_{11}}q_{11}$$

[5] Note from Theorem 3.1 of Devroye (1986, p. 403) that the variable $y_{i1} = q_{i1}^2$,

$$y_{i1} = \frac{x_{i1}^2}{x_{11}^2 + \cdots + x_{n1}^2},$$

has a Beta(1/2,$(n-1)/2$) distribution:

$$p(y_{i1}) = \begin{cases} \frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)} y_{i1}^{(1/2)-1}(1-y_{i1})^{((n-1)/2)-1} & \text{if } y_{i1} \in [0,1] \\ 0 & \text{otherwise} \end{cases}.$$

The density of $q_{i1} = \sqrt{y_{i1}}$ in (32) is then obtained using the change-of-variables formula. Alternatively, one can verify directly that when $\theta \sim U(-\pi, \pi)$, the random variables $\pm\cos\theta$ and $\pm\sin\theta$ each have a density that is given by the special case of (32) corresponding to $n = 2$.

for ω_{11} the $(1,1)$ element of $\mathbf{\Omega}$. In other words, the implicit prior belief about the effect of this structural shock has the distribution of $\sqrt{\omega_{11}}$ times the random variable in (32). Note that if we invoke no further normalization or sign restrictions, "shock number 1" would be associated with no different identifying information than "shock number j". Invariance of the algorithm[6] and the change-of-variables formula for probabilities thus implies a prior density for $h_{ij} = \partial y_{it}/\partial u_{jt}$, the effect on impact of structural shock j on observed variable i given by

$$p(h_{ij}|\mathbf{\Omega}) = \begin{cases} \frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)} \frac{1}{\sqrt{\omega_{ii}}}(1 - h_{ij}^2/\omega_{ii})^{(n-3)/2} & \text{if } h_{ij} \in [-\sqrt{\omega_{ii}}, \sqrt{\omega_{ii}}] \\ 0 & \text{otherwise} \end{cases} . \tag{33}$$

Panel A of Figure 1 plots this density for two different values of n. When there are $n = 2$ variables in the VAR, this prior regards impacts of $\pm\sqrt{\omega_{ii}}$ as more likely than 0. By contrast, when there are $n \geq 4$ variables in the VAR, impacts near 0 are regarded as more likely. The distribution is only uniform in the case $n = 3$.

Alternatively, researchers are often interested in the effect of a structural shock normalized by some definition, for example, the effect on observed variable 2 of a shock that increases observed variable number 1 by one unit. Such an object is characterized by the

[6] This conclusion can also be verified by direct manipulation of the relevant equations, as shown in Appendix E.

ratio of the (2,1) to the (1,1) element of (28):[7]

$$\begin{aligned} h_{21}^* &= \frac{\partial y_{2t}}{\partial u_{1t}^*} = \frac{p_{21}q_{11} + p_{22}q_{21}}{p_{11}q_{11}} = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}}\frac{x_{21}}{x_{11}} \\ &= \omega_{21}/\omega_{11} + \sqrt{\frac{\omega_{22} - \omega_{21}^2/\omega_{11}}{\omega_{11}}}\frac{x_{21}}{x_{11}}. \end{aligned}$$

Recall that the ratio of independent $N(0,1)$ has a standard Cauchy distribution. Invariance properties again establish that the choice of subscripts 1 and 2 is irrelevant. Thus for any $i \neq j$,

$$h_{ij}^*|\mathbf{\Omega} \sim \text{ Cauchy}(c_{ij}^*, \sigma_{ij}^*) \tag{34}$$

with location and scale parameters given by $c_{ij}^* = \omega_{ij}/\omega_{jj}$ and $\sigma_{ij}^* = \sqrt{\frac{\omega_{ii} - \omega_{ij}^2/\omega_{jj}}{\omega_{jj}}}$. This density is plotted in Panel B of Figure 1.

Figure 1 establishes that although the prior is uninformative about the angle of rotation θ, it can be highly informative for the objects about which the researcher intends to form an inference, namely the impulse-response functions.

In practice, applied researchers often ignore posterior uncertainty about $\mathbf{\Omega}$ and simply condition on the average residual variance $\hat{\mathbf{\Omega}}_T$. Identifying normalization or sign restrictions then restrict the distributions $h_{ij}|\hat{\mathbf{\Omega}}_T$ or $h_{ij}^*|\hat{\mathbf{\Omega}}_T$ to certain allowable regions. But within these regions, the shapes of the posterior distributions are exactly those governed by the implicit prior distributions plotted in Figure 1. One can see how this happens using a simple bivariate example in which the first variable is a measure of price and the second

[7] The final equation here is derived from the upper (2×2) block of (29):

$$\begin{bmatrix} \omega_{11} & \omega_{21} \\ \omega_{21} & \omega_{22} \end{bmatrix} = \begin{bmatrix} p_{11}^2 & p_{11}p_{21} \\ p_{11}p_{21} & p_{21}^2 + p_{22}^2 \end{bmatrix}.$$

variable is a measure of quantity. Taking the case of the reflection matrix in (31) for illustration[8] , we have

$$\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} p_{11}\cos\theta & p_{11}\sin\theta \\ (p_{21}\cos\theta + p_{22}\sin\theta) & (p_{21}\sin\theta - p_{22}\cos\theta) \end{bmatrix}. \tag{35}$$

Suppose our normalizing and partially identifying restrictions consist of the requirements that a demand shock raises price ($h_{11} \geq 0$) and raises quantity ($h_{21} \geq 0$) while a supply shock raises price ($h_{12} \geq 0$) and lowers quantity ($h_{22} \leq 0$). The first and third inequalities then restrict $\theta \in [0, \pi/2]$. Suppose further that the correlation between the price and quantity residuals is positive ($p_{21} > 0$). Then $h_{21} \geq 0$ for all $\theta \in [0, \pi/2]$, as called for. But the condition $h_{22} \leq 0$ requires $\theta \in [0, \tilde{\theta}]$ where $\tilde{\theta}$ is the angle in $[0, \pi/2]$ for which $\cot\tilde{\theta} = p_{21}/p_{22}$. Recall that the response of quantity to a supply shock that raises the price by 1% is given by the ratio of the (2,2) to the (2,1) element in (35):

$$h^*_{22} = \frac{p_{21}}{p_{11}} - \frac{p_{22}}{p_{11}}\cot\theta.$$

As θ goes from 0 to $\tilde{\theta}$, h^*_{22} varies from $-\infty$ to 0. One can verify directly[9] that when $\theta \sim U(-\pi, \pi)$, $\cot\theta \sim$ Cauchy(0,1). Thus when the correlation between the OLS residuals is positive, the posterior distribution of h^*_{22} is a Cauchy$(c^*_{22}, \sigma^*_{22})$ truncated to be negative.

The value of h^*_{21} (which measures the response of quantity to a demand shock that raises the price by 1%) is given by

$$h^*_{21} = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}}\tan\theta.$$

[8] Draws using the rotation matrix would always be ruled out by the normalization described below.

[9] See for example Gubner (2006, p. 194). This of course is just a special case of the general result in (34).

As θ varies from 0 to $\tilde{\theta}$, this ranges from

$$h_L = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}} \times 0 = \frac{\omega_{21}}{\omega_{11}}. \tag{36}$$

to

$$h_H = \frac{p_{21}}{p_{11}} + \frac{p_{22}}{p_{11}} \frac{p_{22}}{p_{21}} = \frac{p_{21}^2 + p_{22}^2}{p_{11} p_{21}} = \frac{\omega_{22}}{\omega_{21}} \tag{37}$$

Note that the magnitude h_{21}^* usually goes by another name– it is the short-run price elasticity of supply, while h_{22}^* is just the short-run price elasticity of demand. We have thus seen that, when the correlation between price and quantity is positive, the posterior distribution for the demand elasticity will be the original Cauchy$(c_{22}^*, \sigma_{22}^*)$ truncated to be negative, while the posterior distribution for the supply elasticity will be the original Cauchy$(c_{21}^*, \sigma_{21}^*)$ truncated to the interval $[h_L, h_H]$.

Alternatively, if price and quantity are negatively correlated, the magnitudes in (36) and (37) would be negative numbers. In this case, the posterior distribution for the supply elasticity will be the original Cauchy$(c_{21}^*, \sigma_{21}^*)$ truncated to be positive, while that for the demand elasticity will be the original Cauchy$(c_{22}^*, \sigma_{22}^*)$ truncated to the interval $[h_H, h_L]$.

For illustration, we applied the Rubio-Ramírez, Waggoner, and Zha (2010) algorithm as detailed in Appendix E to an 8-lag VAR fit to U.S. data on growth rates of real labor compensation and total employment over t = 1970:Q1 - 2014:Q2. The reduced-form VAR residual variance matrix is estimated by OLS to be

$$\hat{\mathbf{\Omega}} = \begin{bmatrix} 0.5920 & 0.0250 \\ 0.0250 & 0.1014 \end{bmatrix}. \tag{38}$$

Since the correlation between wages and employment is positive, the set $S(\hat{\mathbf{\Omega}})$ does not restrict the demand elasticity h_{22}^*. The blue histogram in the top panel of Figure 2 plots the magnitude that a researcher using the sign-identification methodology would interpret as the response of employment to a shock to supply that increases the real wage by 1%. The red curve plots a Cauchy$(c_{22}^*, \sigma_{22}^*)$ density truncated to be negative.

The blue histogram in the bottom panel of Figure 2 plots the magnitude that the researcher would describe as the response of employment to a shock to demand that increases the real wage by 1%. From (38) we calculate that $S(\hat{\mathbf{\Omega}})$ restricts this supply elasticity to fall between $h_L = 0.0421$ and $h_H = 4.0626$. The green curve plots a Cauchy$(c_{21}^*, \sigma_{21}^*)$ truncated to be positive and the red curve further truncates it to the interval $[h_L, h_H]$. Because the correlation between the reduced-form residuals is quite small, there is very little difference between the red and green distributions.[10]

The figure illustrates that researchers using the traditional methodology can end up performing hundreds of thousands of calculations, ostensibly analyzing the data, but in the end are doing nothing more than generating draws from a prior distribution that they never even acknowledged that they had assumed!

Another implication of these analytical results is that it would generally be necessary to report the posterior medians rather than posterior means for inference about magnitudes such as the demand elasticity when the traditional method is used because the mean of a

[10] The distributions in Figure 2 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.

Cauchy variable does not exist. We can also see analytically what would happen if we were to apply the traditional methodology to a data set in which the reduced-form variance matrix $\hat{\mathbf{\Omega}}_T$ is diagonal. In this case the posterior distribution for the time-zero impact of any structural shock will be nothing more than the distribution given in (33) truncated either to the positive or negative region.

4 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

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{u}_t'} = \mathbf{\Psi}_s \mathbf{A}^{-1} \tag{39}$$

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

$$\mathbf{F} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_{m-1} & \mathbf{\Phi}_m \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix} \tag{40}$$

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

Beliefs about higher-order impacts are of necessity joint beliefs about $\mathbf{A}$ and $\mathbf{B}$. For example,

$$\frac{\partial \mathbf{y}_t}{\partial \mathbf{u}_t'} = \mathbf{A}^{-1} \tag{41}$$

$$\frac{\partial \mathbf{y}_{t+1}}{\partial \mathbf{u}_t'} = \mathbf{\Phi}_1 \mathbf{A}^{-1}. \tag{42}$$

If $\mathbf{\Phi}_1$ is diagonal with negative elements, the signs of $\partial \mathbf{y}_{t+1}/\partial \mathbf{u}_t'$ are opposite those of $\mathbf{A}^{-1}$ itself. 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 our parameterization prior beliefs about structural impacts for $s > 0$ would be represented in the form of the prior distribution $p(\mathbf{B}|\mathbf{A}, \mathbf{D})$. Our recommendation is that nondogmatic priors should be used for this purpose, since we have seen the data are asymptotically fully informative about the posterior $p(\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T)$.

A common source of the expectation that signs of $\partial \mathbf{y}_{t+s}/\partial \mathbf{u}_t'$ should be the same as those of $\partial \mathbf{y}_t/\partial \mathbf{u}_t'$ is a prior expectation that $\mathbf{\Phi}_1$ is not far from the identity matrix and that elements of $\mathbf{\Phi}_2, ..., \mathbf{\Phi}_m$ are likely small. 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.[11] This suggests that we might want to use priors for $\mathbf{A}$ and $\mathbf{B}$ that imply a value for $\boldsymbol{\eta} = E(\mathbf{\Phi})$ given by

$$\underset{(n\times k)}{\boldsymbol{\eta}} = \begin{bmatrix} \underset{(n\times n)}{\mathbf{I}_n} & \underset{[n\times(k-n)]}{\mathbf{0}} \end{bmatrix}. \tag{43}$$

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

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

$\mathbf{B}|\mathbf{A}$ to be

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

suggesting a prior mean for $\mathbf{b}_i$ given by $\tilde{\mathbf{m}}_i = E(\mathbf{b}_i|\mathbf{A}) = \boldsymbol{\eta}'\mathbf{a}_i$. We can also follow Doan, Litterman and Sims (1984) as modified by Sims and Zha (1998) in putting more confidence in our prior beliefs that higher-order lags are zero, as we describe in detail in Appendix D.

Some researchers may want to use additional prior information about structural dynamics. In the example in the following section, we consider a prior belief that a labor demand shock should have little permanent effect on the level of employment. We have found it convenient to implement these as supplements to the general recommendations in Appendix D, where we could put as little weight as we wanted on the general recommendations by specifying λ_0 in expression (64) to be sufficiently large. For example, suppose we wanted to supplement the beliefs about reduced-form coefficients from the Doan, Litterman and Sims prior with the additional prior belief that h_i linear combinations (represented by $\mathbf{R}_i\mathbf{b}_i$) should be close to some expected value $\mathbf{r}_i$, where $\mathbf{r}_i$ could be a function of $\mathbf{A}$. As in Theil (1971, pp. 347-49) it is convenient to represent this additional information in the form of h_i pseudo observations

$$\underset{(h_i\times 1)}{\mathbf{r}_i} = \underset{(h_i\times k)}{\mathbf{R}_i}\underset{(k\times 1)}{\mathbf{b}_i} + \underset{(h_i\times 1)}{\mathbf{v}_i} \qquad \mathbf{v}_i \sim N(\mathbf{0}, d_{ii}\underset{(h_i\times h_i)}{\mathbf{V}_i}) \tag{44}$$

allowing us to simply replace (13) and (14) with

$$\underset{[(T+k+h_i)\times 1]}{\tilde{\mathbf{Y}}_i} = \begin{bmatrix} \mathbf{a}_i'\mathbf{y}_1 & \cdots & \mathbf{a}_i'\mathbf{y}_T & \mathbf{m}_i'\mathbf{P}_i & \mathbf{r}_i'\mathbf{P}_{V_i} \end{bmatrix}' \tag{45}$$

$$\underset{[(T+k+h_i)\times k]}{\tilde{\mathbf{X}}_i} = \begin{bmatrix} \mathbf{x}_0 & \cdots & \mathbf{x}_{T-1} & \mathbf{P}_i & \mathbf{R}_i'\mathbf{P}_{V_i} \end{bmatrix}' \tag{46}$$

for $\mathbf{M}_i^{-1} = \mathbf{P}_i\mathbf{P}_i'$ and $\mathbf{V}_i^{-1} = \mathbf{P}_{V_i}\mathbf{P}_{V_i}'$. This is numerically identical to using $\mathbf{b}_i|\mathbf{A},\mathbf{D} \sim N(\tilde{\mathbf{m}}_i, d_{ii}\tilde{\mathbf{M}}_i)$ as a unified prior in (10) with $\tilde{\mathbf{M}}_i^{-1} = \mathbf{M}_i^{-1}+\mathbf{R}_i'\mathbf{V}_i^{-1}\mathbf{R}_i$ and $\tilde{\mathbf{m}}_i = \tilde{\mathbf{M}}_i(\mathbf{M}_i^{-1}\mathbf{m}_i+\mathbf{R}_i'\mathbf{V}_i^{-1}\mathbf{r}_i)$.

Although such priors can help improve inference about the structural parameters in a given observed sample, they do not change any of the asymptotics in Proposition 2.

5 Application: Bayesian inference in a model of labor supply and demand.

In this section we illustrate these methods using the example of labor supply and labor demand introduced in Section 3. We noted in Figure 2 how prior beliefs influence the results when the traditional sign-restriction methodology is applied to these data. The goal of this section is to elevate such prior information from something that the researcher imposes mechanically without consideration to something that is explicitly acknowledged and can be motivated from economic theory and empirical evidence from other data sets. In order to do this, the first step is to write down the structural model with which we are proposing to interpret the observed correlations. Consider dynamic labor demand and supply curves of the form:

$$\begin{aligned}
\text{demand:} \qquad & \Delta n_t = k^d + \beta^d \Delta w_t + b_{11}^d \Delta w_{t-1} + b_{12}^d \Delta n_{t-1} + b_{21}^d \Delta w_{t-2} + b_{22}^d \Delta n_{t-2} + \\
& \cdots + b_{m1}^d \Delta w_{t-m} + b_{m2}^d \Delta n_{t-m} + u_t^d \\
\\
\text{supply:} \qquad & \Delta n_t = k^s + \alpha^s \Delta w_t + b_{11}^s \Delta w_{t-1} + b_{12}^s \Delta n_{t-1} + b_{21}^s \Delta w_{t-2} + b_{22}^s \Delta n_{t-2} + \\
& \cdots + b_{m1}^s \Delta w_{t-m} + b_{m2}^s \Delta n_{t-m} + u_t^s.
\end{aligned} \tag{47}$$

Here Δn_t is the growth rate of total U.S. employment,[12] Δw_t is the growth rate of real compensation per hour,[13] β^d is the short-run wage elasticity of demand, and α^s is the short-run wage elasticity of supply. Note that the system (47) is a special case of (1) with $\mathbf{y}_t = (\Delta w_t, \Delta n_t)'$ and

$$\mathbf{A} = \begin{bmatrix} -\beta^d & 1 \\ -\alpha^s & 1 \end{bmatrix}. \tag{48}$$

OLS estimation of the reduced-form (2) for this system with $m = 8$ lags and $t =$ 1970:Q1 through 2014:Q2 led to the estimate of the reduced-form residual variance matrix $\hat{\mathbf{\Omega}} = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$ reported in (38).

5.1 Maximum likelihood estimates.

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

$$\hat{\beta}(\alpha) = \frac{\sum_{t=1}^{T}(\hat{\varepsilon}_{2t} - \alpha\hat{\varepsilon}_{1t})\hat{\varepsilon}_{2t}}{\sum_{t=1}^{T}(\hat{\varepsilon}_{2t} - \alpha\hat{\varepsilon}_{1t})\hat{\varepsilon}_{1t}} = \frac{(\hat{\omega}_{22} - \alpha\hat{\omega}_{12})}{(\hat{\omega}_{12} - \alpha\hat{\omega}_{11})}, \tag{49}$$

for $\hat{\omega}_{ij}$ the (i,j) element of $\hat{\mathbf{\Omega}}$. One can verify directly that any pair (α, β) satisfying (49) produces a diagonal matrix for $\mathbf{A}\hat{\mathbf{\Omega}}\mathbf{A}'$.

The top panel of Figure 3 plots the function $\hat{\beta}(\alpha)$ for these data. Any pair (α, β) lying

[12] The level n_t was measured as 100 times the natural log of the seasonally adjusted number of people on nonfarm payrolls during the third month of quarter t from series PAYEMS downloaded August 2014 from http://research.stlouisfed.org/fred2/.

[13] The level w_t was measured as 100 times the natural log of seasonally adjusted real compensation per hour for the nonfarm business sector from series COMPRNFB downloaded August 2014 from http://research.stlouisfed.org/fred2/.

on these curves would maximize the likelihood function, and there is no basis in the data for preferring one point on the curves 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 lower right quadrant in the figure. When, as in this data set, the OLS residuals are positively correlated, the sign restrictions are consistent with any $\beta \in (-\infty, 0)$, but require α to fall in the interval (h_L, h_H) defined in (36)-(37). We earlier derived these bounds considering allowable angles of rotation but it is also instructive to explain their intuition in terms of a structural interpretation of the likelihood function.[14] Note that h_L 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 h_L and the MLE for α can be no smaller than h_L. The fact that the MLE for β can be no larger than h_L is not a restriction, because we have separately required that $\beta < 0$ and in the case under discussion, $h_L > 0$. However, the inference that α can be no smaller than h_L puts a lower bound on α. At the other end, the OLS coefficient from a regression of $\hat{\varepsilon}_{1t}$ on $\hat{\varepsilon}_{2t}$ (that is, h_H^{-1}) turns out to be a weighted average of α^{-1} and β^{-1}, requiring $\beta^{-1} < h_H^{-1}$ (again not binding when $h_H > 0$) and $\alpha^{-1} > h_H^{-1}$; the latter gives us the upper bound that $\alpha < h_H$. This is the intuition for why $h_L < \alpha < h_H$.

The bottom panel in Figure 3 plots contours of the concentrated likelihood function, that

[14] Leamer (1981) discovered these points in a simple OLS setting years ago.

is, contours of the function

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

The data are quite 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 $\beta = \hat{\beta}(\alpha)$ shown in black.

The set $S(\mathbf{\Omega})$ in expression (23) is calculated for this example as follows. When the correlation between the VAR residuals ω_{12} is positive, $S(\mathbf{\Omega})$ is the set of all $\mathbf{A}$ in (48) such that $\beta \leq 0$, $(\omega_{21}/\omega_{11}) \leq \alpha \leq (\omega_{22}/\omega_{21})$, and $\beta = (\omega_{22} - \alpha\omega_{12})/(\omega_{12} - \alpha\omega_{11})$, in other words, the set of points on the black curve in Figure 3 between h_L and h_H.

5.2 Prior information about elasticities.

While the frequentist would regard any of the points in $S(\hat{\mathbf{\Omega}})$ as equally plausible, from the standpoint of economic theory that is surely not the case; for example, a demand elasticity of -100 would not be consistent with any coherent economic model. We propose to represent prior information about α and β using Student t distributions with ν degrees of freedom. Note that this includes the Cauchy distribution in (34) as a special case when $\nu = 1$. One benefit of using $\nu \geq 3$ is that the posterior distributions for α and β would then be guaranteed to have finite mean and variance. Our proposal is that the location and scale parameters for these distributions should be chosen on the basis of prior information about these elasticities.

Hamermesh's (1996) survey of microeconometric studies concluded that the absolute value of the elasticity of labor demand is between 0.15 and 0.75. Lichter, Peichl, and Siegloch's (2014) meta-analysis of 942 estimates from 105 different studies favored values at the lower end of this range. On the other hand, theoretical macro models can imply a value of 2.5 or higher (Akerlof and Dickens, 2007; Galí, Smets, and Wouters, 2012). A prior for β that reflects the uncertainty associated with these findings could be represented with a Student t distribution with location parameter $c_\beta = -0.6$, scale parameter $\sigma_\beta = 0.6$, and degrees of freedom $\nu_\beta = 3$, truncated to be negative:

$$p(\beta) = \begin{cases} F(0; c_\beta, \sigma_\beta, \nu_\beta)^{-1} f(\beta; c_\beta, \sigma_\beta, \nu_\beta) & \text{if } \beta \leq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{50}$$

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,

$$f(x; c, \sigma, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\sigma\Gamma(\nu/2)} \left(1 + \frac{(x-c)^2}{\sigma^2\nu}\right)^{-(\nu+1)/2} \tag{51}$$

and $F(.)$ the cumulative distribution function $F(x; c, \sigma, \nu) = \int_{-\infty}^{x} f(z; c, \sigma, \nu) dz$. The prior in (50) places a 5% probability on values of $\beta < -2.2$ and another 5% probability on values above -0.1.

In terms of the labor supply elasticity, a key question is whether the increase in wages is viewed as temporary or permanent. A typical assumption is that the income and substitution effects cancel, in which case there would be zero observed response of labor supply to a permanent increase in the real wage (Kydland and Prescott, 1982). On the other hand, the response to a temporary wage increase is often interpreted as the Frisch (or compensated marginal utility) elasticity, about which there are again quite different consensuses in the micro and macro literatures. Chetty et al. (2013) reviewed 15 different quasi-experimental studies all of which implied Frisch elasticities below 0.5. A separate survey of microeconometric studies by Reichling and Whalen (2012) concluded that the Frisch elasticity is between 0.27 and 0.53. By contrast, values above one or two are common in the macroeconomic literature (see for example Kydland and Prescott, 1982, Cho and Cooley, 1994, and Smets and Wouters, 2007). For the prior used in this study, we specified $c_\alpha = 0.6$, $\sigma_\alpha = 0.6$, and $\nu_\alpha = 3$, which associates a 90% probability with $\alpha \in (0.1, 2.2)$:

$$p(\alpha) = \begin{cases} [1 - F(0; c_\alpha, \sigma_\alpha, \nu_\alpha)]^{-1} f(\alpha; c_\alpha, \sigma_\alpha, \nu_\alpha) & \text{if } \alpha \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

The prior distribution $p(\mathbf{A})$ was thus taken to be the product of (50) with (52). Contours for this prior distribution are provided in the top panel of Figure 4, while the bottom panel displays contours for the posterior distribution that would result if we used only this prior $p(\mathbf{A})$ with no additional information about $\mathbf{D}$ or $\mathbf{B}$. The observed data contain sufficient information to cause all of the posterior distribution to fall within a close neighborhood of $S(\hat{\mathbf{\Omega}})$. But whereas the frequentist regards all points within this set as equally plausible, a sensible person with knowledge of the literature would regard values such as $\alpha = 0.5$, $\beta = -0.3$ as much more plausible than $\alpha = 0.06$, $\beta = -9.5$. The Bayesian approach gives the analyst a single coherent framework for combining uncertainty caused by observing a limited data set (that is, uncertainty about the true value of $\mathbf{\Omega}$) with uncertainty about the correct structure of the model itself (that is, uncertainty about points within $S(\mathbf{\Omega})$), which combined uncertainty is represented by the contours in the bottom panel of Figure 4.

5.3 Long-run restrictions.

Unfortunately, this combined uncertainty remains quite large for this example; we have no basis for distinguishing alternative elements of $S(\hat{\mathbf{\Omega}})$ other than the priors just described. To infer anything more from the data would require imposing additional structure, which from a Bayesian perspective means drawing on additional sources of outside information. We illustrate how this could be done by making use of additional beliefs about the long-run labor supply elasticity. Let $\tilde{\mathbf{y}}_t = (w_t, n_t)'$ denote the levels of the variables so that $\mathbf{y}_t = \Delta\tilde{\mathbf{y}}_t$. From (39) the effect of the structural shocks on the future levels of wages and employment is given by

$$\begin{aligned} \frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} &= \frac{\partial \Delta\tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} + \frac{\partial \Delta\tilde{\mathbf{y}}_{t+s-1}}{\partial \mathbf{u}_t'} + \cdots + \frac{\partial \Delta\tilde{\mathbf{y}}_{t}}{\partial \mathbf{u}_t'} \\ &= \mathbf{\Psi}_s\mathbf{A}^{-1} + \mathbf{\Psi}_{s-1}\mathbf{A}^{-1} + \cdots + \mathbf{\Psi}_0\mathbf{A}^{-1} \end{aligned} \quad (53)$$

with a permanent or long-run effect given by

$$\begin{aligned} \lim_{s\to\infty} \frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} &= (\mathbf{\Psi}_0 + \mathbf{\Psi}_1 + \mathbf{\Psi}_2 + \cdots)\mathbf{A}^{-1} \\ &= (\mathbf{I}_n - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_m)^{-1}\mathbf{A}^{-1} \\ &= [\mathbf{A}(\mathbf{I}_n - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_m)]^{-1}. \end{aligned} \quad (54)$$

The long-run effect of a labor demand shock on employment is given by the (2,1) element of this matrix. This long-run elasticity would be zero if and only if the above matrix is upper triangular, or equivalently if and only if the following matrix is upper triangular

$$\begin{aligned} \mathbf{A}(\mathbf{I}_n - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_m) &= \mathbf{A}(\mathbf{I}_n - \mathbf{A}^{-1}\mathbf{B}_1 - \mathbf{A}^{-1}\mathbf{B}_2 - \cdots - \mathbf{A}^{-1}\mathbf{B}_m) \\ &= \mathbf{A} - \mathbf{B}_1 - \mathbf{B}_2 - \cdots - \mathbf{B}_m \end{aligned} \quad (55)$$

requiring

$$0 = -\alpha^s - b_{11}^s - b_{21}^s - \cdots - b_{m1}^s$$

$$b_{11}^s + b_{21}^s + \cdots + b_{m1}^s = -\alpha^s. \tag{56}$$

If we were to insist that (56) has to hold exactly, then the model would become just-identified even in the absence of any information about α^s or β^d, and indeed hundreds of empirical papers have used exactly such a procedure to perform structural inference using VARs.[15] We propose instead to represent the idea as a prior belief of the form

$$(b_{11}^s + b_{21}^s + \cdots + b_{m1}^s)|\mathbf{A}, \mathbf{D} \sim N(-\alpha^s, d_{22}V_2). \tag{57}$$

Note that our proposal is a strict generalization of the existing approach, in that (57) becomes (56) in the special case when $V_2 \to 0$. We would further argue that our method is a strict improvement over the existing approach in several respects. First, (56) is usually implemented by conditioning on the reduced-form estimates $\hat{\mathbf{\Phi}}$. By contrast, our approach will generate the statistically optimal joint inference about $\mathbf{A}$ and $\mathbf{B}$ taking into account the fact that both are being estimated with error. Second, we would argue that the claim that (56) is known with certainty is indefensible. A much better approach in our view is to acknowledge openly that (56) is a prior belief about which any reasonable person would have some uncertainty. Granted, an implication of Proposition 2 is that some of this uncertainty will necessarily remain even if we had available an infinite sample of observations on $\mathbf{y}$. However, our position is that this uncertainty about the specification should be openly

[15] See for example Shapiro and Watson (1988), Blanchard and Quah (1989), and Galí (1999).

acknowledged and reported as part of the results, and indeed as we have demonstrated this is exactly what is accomplished using the algorithm suggested in Proposition 1.

In addition to these prior beliefs about the long-run elasticity, we also used the general priors suggested in Appendix D, using (13) and (14) for $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{X}}_1$ and (45) and (46) for $\tilde{\mathbf{Y}}_2$ and $\tilde{\mathbf{X}}_2$ with $r_2 = -\alpha^s$, $V_2 = 0.1$, and

$$\underset{(1\times k)}{\mathbf{R}_2} = \begin{bmatrix} \mathbf{1}'_m \otimes \mathbf{e}'_2 & 0 \end{bmatrix}$$

for $\mathbf{1}_m$ an $(m \times 1)$ vector of ones and $\mathbf{e}'_2 = (1, 0)$. The value for V_2 corresponds to putting a weight on the long-run restriction equivalent to 10 observations.

Note that our recommended approach also helps address some of the concerns about long-run restrictions raised by Faust and Leeper (1997). For example, rather than dogmatically impose that coefficients beyond some fixed lag are all zero, our approach instead allows the researcher to shrink coefficients toward zero gradually as the lag length increases up to some large m, with the pace of the shrinkage governed by choice of the parameter λ_1 in expression (63).

5.4 Empirical results.

The top panels in Figure 5 display prior densities (red curves) and posterior densities (blue histograms) for the short-run demand and supply elasticities. The data cause us to revise our prior beliefs about β^d, regarding elasticities near zero as less likely having seen these data. Our beliefs about the short-run supply elasticity are more strongly revised, favoring estimates at the lower range of the microeconometric literature over values often assumed in macroeconomic studies. Although our prior expectation was for a zero long-run response of employment to a labor demand shock, the data provide support for a significant positive permanent effect (see the bottom panel in Figure 5).

Median posterior values for the impulse-response functions in (53) are plotted as the solid lines in Figure 6. The shaded 95% posterior credible regions reflect both uncertainty associated with having observed only a finite set of data as well as uncertainty about the true structure. A 1% leftward shift of the labor demand curve raises worker compensation on impact (and permanently as well) by about 1%, and raises employment on impact by much less than 1% in equilibrium due to the limited short-run labor-supply elasticity. And although we approached the data with an expectation that this would not have a permanent effect on employment, after seeing the data we would be persuaded that it does. An increase in the number of people looking for work depresses labor compensation (upper right panel of Figure 6), and raises employment over time.

Figure 7 shows the consequences of putting different weights on the prior belief about the long-run labor-supply elasticity. The first panel in the second row reproduces the impulse-response function from the lower-left panel of Figure 6, while the second panel in that row reproduces the prior and posterior distributions for the short-run labor supply elasticity (the upper-right panel of Figure 5, drawn here on a different scale for easier visual comparisons). The first row of Figure 7 shows the effects of a weaker weight on the long-run restrictions, $V_2 = 1.0$, weighting the long-run belief as equivalent to only one observation rather than 10. With weaker confidence about the long-run effect, we would not conclude that the short-run labor supply elasticity was so low or that the equilibrium effects of a labor demand shock on employment were as muted. The third and fourth rows show the effects of using more informative priors ($V_2 = 0.01$ and $V_2 = 0.001$, respectively). Even when we have a fairly tight prior represented by $V_2 = 0.01$, the data still lead us away from believing that the long-run effect of a labor demand shock is literally zero, and to reach that conclusion we need to impute a very small value to the short-run labor supply elasticity as well. Only when we specify $V_2 = 0.001$ is the long-run effect pushed all the way to zero.

This exercise demonstrates an important advantage of representing prior information in the form proposed in Section 2. In a typical frequentist approach, the restriction (56) is viewed as necessary to arrive at a just-identified model and is therefore regarded as inherently untestable. By contrast, we have seen that if we instead regard it as one of a set of nondogmatic prior beliefs, it is possible to examine what role the assumption plays in determining the final results and to assess its plausibility. Our conclusion from this exercise is that it would not be a good idea to rely on exact satisfaction of (56) as the identifying assumption for structural analysis. A combination of a weaker belief in the long-run impact along with information from other sources about short-run impacts is a superior approach.

6 Conclusion.

Drawing structural inference from observed correlations requires making use of prior beliefs about economic structure. In just-identified models, researchers usually proceed as if these prior beliefs are known with certainty. In vector autoregressions that are only partially identified using sign restrictions, the way that implicit prior beliefs influence the reported results has not been recognized in the previous literature. In this paper we have explicated the prior beliefs that are implicit in sign-restricted VARs and proposed a general Bayesian framework that can be used to make optimal use of prior information and elucidate the consequences of prior beliefs in any vector autoregression. Our suggestion is that explicitly defending the prior information used in the analysis and reporting the way in which the observed data causes these prior beliefs to be revised is superior to pretending that prior information was not used and has no effect on the reported conclusions.

Appendix

A. Proof of Proposition 1.

The likelihood (12) can be written

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

If we define $\mathbf{X}_i^* = \mathbf{P}_i'$ and $\mathbf{y}_i^* = \mathbf{P}_i'\mathbf{m}_i$, the prior for $\mathbf{b}_i$ in (10) 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 $\tilde{\mathbf{X}}_i$ as represented by the value of $\mathbf{m}_i^*$ in equation (15). 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

$$\begin{aligned}
(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{b}_i)'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{b}_i) &= (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^* + \tilde{\mathbf{X}}_i\mathbf{m}_i^* - \tilde{\mathbf{X}}_i\mathbf{b}_i)'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^* + \tilde{\mathbf{X}}_i\mathbf{m}_i^* - \tilde{\mathbf{X}}_i\mathbf{b}_i) \\
&= (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^*)'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\mathbf{m}_i^*) + (\mathbf{b}_i - \mathbf{m}_i^*)'\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i(\mathbf{b}_i - \mathbf{m}_i^*) \\
&= \tilde{\mathbf{Y}}_i'\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}_i'\tilde{\mathbf{X}}_i(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i + (\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*) \\
&= \zeta_i^* + (\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*).
\end{aligned}$$

The product of the likelihood (12) with the prior for $\mathbf{B}$ (9) can thus be written

$$p(\mathbf{B}|\mathbf{A},\mathbf{D})p(\mathbf{Y}_T|\mathbf{A},\mathbf{D},\mathbf{B}) = (2\pi)^{-Tn/2}|\det(\mathbf{A})|^T|\mathbf{D}|^{-T/2}\times \tag{58}$$
$$\prod_{i=1}^n \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i|^{1/2}}\exp\left[-\frac{\zeta_i^* + (\mathbf{b}_i - \mathbf{m}_i^*)'(\mathbf{M}_i^*)^{-1}(\mathbf{b}_i - \mathbf{m}_i^*)}{2d_{ii}}\right].$$

Multiplying (58) by the priors for $\mathbf{A}$ and $\mathbf{D}$ and rearranging gives

$$\begin{aligned}
p(\mathbf{Y}_T, \mathbf{A}, \mathbf{D}, \mathbf{B}) &= p(\mathbf{A})p(\mathbf{D}|\mathbf{A})p(\mathbf{B}|\mathbf{A}, \mathbf{D})p(\mathbf{Y}_T|\mathbf{A}, \mathbf{D}, \mathbf{B}) \\
&= p(\mathbf{A})(2\pi)^{-Tn/2}|\det(\mathbf{A})|^T \prod_{i=1}^{n} \left\{ d_{ii}^{-T/2} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)} \frac{\Gamma(\kappa_i^*)}{(\tau_i^*)^{\kappa_i^*}} \frac{(\tau_i^*)^{\kappa_i^*}}{\Gamma(\kappa_i^*)} (d_{ii}^{-1})^{\kappa_i - 1} \exp(-\tau_i^* d_{ii}^{-1}) \times \right. \\
&\quad \left. \frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}} \frac{1}{(2\pi)^{k/2}|d_{ii}\mathbf{M}_i^*|^{1/2}} \exp\left[-\frac{(\mathbf{b}_i - \mathbf{m}_i^*)' (\mathbf{M}_i^*)^{-1} (\mathbf{b}_i - \mathbf{m}_i^*)}{2d_{ii}} \right] \right\} \\
&= p(\mathbf{A})(2\pi)^{-Tn/2}|\det(\mathbf{A})|^T \times \\
&\quad \prod_{i=1}^{n} \left\{ \frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)} \frac{\Gamma(\kappa_i^*)}{(\tau_i^*)^{\kappa_i^*}} \right\} \gamma(d_{ii}^{-1}; \kappa_i^*, \tau_i^*) \phi(\mathbf{b}_i; \mathbf{m}_i^*, d_{ii}\mathbf{M}_i^*). \qquad (59)
\end{aligned}$$

Note that the product in (59) 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

$$p(\mathbf{Y}_T)p(\mathbf{A}|\mathbf{Y}_T) = p(\mathbf{A})(2\pi)^{-Tn/2}|\det(\mathbf{A})|^T \prod_{i=1}^{n} \left\{ \frac{|\mathbf{M}_i^*|^{1/2}}{|\mathbf{M}_i|^{1/2}} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)} \frac{\Gamma(\kappa_i^*)}{(\tau_i^*)^{\kappa_i^*}} \right\}. \qquad (60)$$

The posterior $p(\mathbf{A}|\mathbf{Y}_T)$ is thus proportional to (60). Since $\hat{\mathbf{\Omega}}_T$ is not a function of $\mathbf{A}$, we can write multiply (60) by $|\hat{\mathbf{\Omega}}_T|^{T/2}$ to write the result in an equivalent form that facilitates numerical calculation and interpretation:

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

as claimed in equation (20). Note that we could replace $\hat{\mathbf{\Omega}}_T$ in the numerator with any matrix not depending on unknown parameters, with any such replacement simply changing the definition of k_T in (20). 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 $\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T$. Define the target function to be

$$\begin{aligned} q(\mathbf{A}) &= \log p(\mathbf{A}) + (T/2)\log[\det(\mathbf{A}\hat{\boldsymbol{\Omega}}_T\mathbf{A}')] \\ &\quad - \sum_{i=1}^{n}(\kappa_i(\mathbf{A}) + T/2)\log\{[2\tau_i(\mathbf{A})/T] + [\zeta_i^*(\mathbf{A})/T]\} + \sum_{i=1}^{n}\kappa_i(\mathbf{A})\log\tau_i(\mathbf{A}) \\ &\quad + \sum_{i=1}^{n}\{(1/2)\log[|\mathbf{M}_i^*(\mathbf{A})|/|\mathbf{M}_i(\mathbf{A})|] + \log[\Gamma(\kappa_i(\mathbf{A}) + T/2)/\Gamma(\kappa_i(\mathbf{A})]\}. \end{aligned} \tag{61}$$

If κ_i and $\mathbf{M}_i$ do not depend on $\mathbf{A}$ (as in the priors recommended in Appendix D) the last term in (61) can be dropped. Those priors specify $\tau_i(\mathbf{A}) = \kappa_i\mathbf{a}_i'\mathbf{S}\mathbf{a}_i$ and $\zeta_i^*(\mathbf{A}) = \mathbf{a}_i'\mathbf{W}\mathbf{a}_i$ where

$$\begin{aligned} \mathbf{W} &= \sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' + \boldsymbol{\eta}\mathbf{M}^{-1}\boldsymbol{\eta}' - \\ &\quad \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}' + \boldsymbol{\eta}\mathbf{M}^{-1}\right)\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}' + \mathbf{M}^{-1}\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{y}_t' + \mathbf{M}^{-1}\boldsymbol{\eta}'\right) \end{aligned}$$

for $\mathbf{M}$ the diagonal matrix given in (65) and $\boldsymbol{\eta}$ the matrix in (43). For more general priors, $\zeta_i^*(\mathbf{A})$ can be calculated from (19).

As a first step we calculate an approximation to the shape of the posterior distribution that will help the numerical components of the algorithm be more efficient. Collect elements

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

of $\mathbf{A}$ that are not known with certainty in an $(n_\alpha \times 1)$ vector $\boldsymbol{\alpha}$, and find the value $\hat{\boldsymbol{\alpha}}$ that maximizes (61) numerically. This value $\hat{\boldsymbol{\alpha}}$ offers a reasonable guess for the posterior mean of $\boldsymbol{\alpha}$, while the matrix of second derivatives (again obtained numerically) gives an idea of the scale of the posterior distribution:

$$\hat{\boldsymbol{\Lambda}} = \left. \frac{\partial^2 q(\mathbf{A}(\boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} .$$

We then use this guess to inform a random-walk Metropolis Hastings algorithm to generate candidate draws of $\boldsymbol{\alpha}$ from the posterior distribution, as follows. We can begin the algorithm at step 1 by setting $\boldsymbol{\alpha}^{(1)} = \hat{\boldsymbol{\alpha}}$. As a result of step ℓ we have generated a value of $\boldsymbol{\alpha}^{(\ell)}$. For step $\ell + 1$ we generate

$$\tilde{\boldsymbol{\alpha}}^{(\ell+1)} = \boldsymbol{\alpha}^{(\ell)} + \xi \left(\hat{\mathbf{P}}_{\boldsymbol{\Lambda}}^{-1} \right)' \mathbf{v}_t$$

for $\mathbf{v}_t$ an $(n_\alpha \times 1)$ vector of Student t variables with 2 degrees of freedom, $\hat{\mathbf{P}}_{\boldsymbol{\Lambda}}$ the Cholesky factor of $\hat{\boldsymbol{\Lambda}}$ (namely $\hat{\mathbf{P}}_{\boldsymbol{\Lambda}} \hat{\mathbf{P}}'_{\boldsymbol{\Lambda}} = \hat{\boldsymbol{\Lambda}}$ with $\hat{\mathbf{P}}_{\boldsymbol{\Lambda}}$ lower triangular), and ξ a tuning scalar to be described shortly. If $q(\mathbf{A}(\tilde{\boldsymbol{\alpha}}^{(\ell+1)})) < q(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)}))$, we set $\boldsymbol{\alpha}^{(\ell+1)} = \boldsymbol{\alpha}^{(\ell)}$ with probability $1 - \exp \left[q(\mathbf{A}(\tilde{\boldsymbol{\alpha}}^{(\ell+1)})) - q(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})) \right]$; otherwise, we set $\boldsymbol{\alpha}^{(\ell+1)} = \tilde{\boldsymbol{\alpha}}^{(\ell+1)}$. The parameter ξ is chosen so that about 30% of the newly generated $\tilde{\boldsymbol{\alpha}}^{(\ell+1)}$ get retained. The values after the first D burn-in draws $\{\boldsymbol{\alpha}^{(D+1)}, \boldsymbol{\alpha}^{(D+2)}, ..., \boldsymbol{\alpha}^{(D+N)}\}$ represent a sample of size N drawn from the posterior distribution $p(\boldsymbol{\alpha}, \mathbf{D}, \mathbf{B} | \mathbf{Y}_T)$; in our applications we have used $D = N = 10^6$.

For each of these N final values for $\boldsymbol{\alpha}^{(\ell)}$ we further generate $\delta_{ii}^{(\ell)} \sim \Gamma(\kappa_i^*(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})), \tau_i^*(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})))$ for $i = 1, ..., n$ and take $\mathbf{D}^{(\ell)}$ to be a diagonal matrix whose row i, column i element is given by $1/\delta_{ii}^{(\ell)}$. From these we also generate $\mathbf{b}_i^{(\ell)} \sim N(\mathbf{m}_i^*(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})), d_{ii}^{(\ell)} \mathbf{M}_i^*(\mathbf{A}(\boldsymbol{\alpha}^{(\ell)})))$ for $i = 1, ..., n$

and take $\mathbf{B}^{(\ell)}$ the matrix whose ith row is given by $\mathbf{b}_i^{(\ell)\prime}$. The triple $\{\mathbf{A}(\boldsymbol{\alpha}^{(\ell)}), \mathbf{D}^{(\ell)}, \mathbf{B}^{(\ell)}\}_{\ell=D+1}^{D+N}$ then represents a sample of size N drawn from the posterior distribution $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$.

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

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

Hence

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

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

$$
\begin{aligned}
(iii) \quad E[(d_{ii}^{-1} - \kappa_i^*/\tau_i^*)^2|\mathbf{Y}_T] &= \kappa_i^*/\tau_i^{*2} \\
&= \frac{\kappa_i + (T/2)}{[\tau_i + (\zeta_i^*/2)]^2} \\
&= \frac{(\kappa_i/T) + (1/2)}{T\,[(\tau_i/T) + (\zeta_i^*/2T)]^2} \\
&\xrightarrow{p} 0
\end{aligned}
$$

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

$$
\text{Prob}\{[\mathbf{A} \notin \mathbf{H}_\delta(\hat{\mathbf{\Omega}}_T)]|\mathbf{Y}_T\} \to 0 \qquad \forall \delta > 0. \tag{62}
$$

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 $\mathbf{P}(\mathbf{A}, \mathbf{\Omega})[\mathbf{P}(\mathbf{A}, \mathbf{\Omega})]' = \mathbf{A\Omega A}'$. Note that

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

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

Furthermore, ζ_i^*, the sum of squared residuals from a regression of $\tilde{\mathbf{Y}}_i$ on $\tilde{\mathbf{X}}_i$, by construction is larger than $T\mathbf{a}_i'\hat{\mathbf{\Omega}}_T\mathbf{a}_i$, the SSR from a regression of $\mathbf{a}_i'\mathbf{y}_t$ on $\mathbf{x}_{t-1}$. Thus

$$
\begin{aligned}
(2\tau_i^*/T) &= (2\tau_i/T) + \zeta_i^*/T \\
&\geq \mathbf{a}_i'\hat{\mathbf{\Omega}}_T\mathbf{a}_i \\
&= \textstyle\sum_{j=1}^i [p_{ij}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)]^2.
\end{aligned}
$$

But for all $\mathbf{A} \notin H_\delta(\hat{\mathbf{\Omega}}_T)$, $\exists\, j^* < i^*$ such that $\left[p_{i^*j^*}(\mathbf{A}, \hat{\mathbf{\Omega}}_T)\right]^2 > \delta^*$ for $\delta^* = 2\delta/[n(n-1)]$ meaning $\mathbf{a}'_{i^*}\hat{\mathbf{\Omega}}_T\mathbf{a}_{i^*} > [\delta^* + p_{i^*i^*}(\mathbf{A}, \hat{\mathbf{\Omega}}_T)]^2$ for some i^* and $\prod_{i=1}^n (2\tau_i^*/T) > [p_{11}(\mathbf{A}, \hat{\mathbf{\Omega}}_T]^2 \cdots [\delta^* + p_{i^*i^*}(\mathbf{A}, \hat{\mathbf{\Omega}}_T)]^2 \cdots [p_{nn}(\mathbf{A}, \hat{\mathbf{\Omega}}_T]^2$. Thus

$$\mathrm{Prob}[\mathbf{A} \notin H_\delta(\hat{\mathbf{\Omega}}_T)|\mathbf{Y}_T] = \int_{\mathbf{A}\notin H_\delta(\hat{\mathbf{\Omega}}_T)} \frac{k_T p(\mathbf{A})[p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)p_{22}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T) \cdots p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)]^{T/2}}{\prod_{i=1}^n [(2\tau_i/T) + \boldsymbol{\zeta}_i^*/T]^{\kappa_i}\,[(2\tau_i/T) + \zeta_i^*/T]^{T/2}} d\mathbf{A}$$

$$\begin{aligned} &< \int_{\mathbf{A}\notin H_\delta(\hat{\mathbf{\Omega}}_T)} \left[\frac{k_T p(\mathbf{A})}{\prod_{i=1}^n [(2\tau_i/T) + (\boldsymbol{\zeta}_i^*/T)]^{\kappa_i}} \times \right. \\ &\qquad \left. \frac{[p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)p_{22}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T) \cdots p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)]^{T/2}}{\{[p_{11}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)] \cdots [\delta^* + p_{i^*i^*}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)] \cdots [p_{nn}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T]\}^{T/2}}\right] d\mathbf{A} \\ &= \int_{\mathbf{A}\notin H_\delta(\hat{\mathbf{\Omega}}_T)} \frac{k_T p(\mathbf{A})}{\prod_{i=1}^n [(2\tau_i/T) + \zeta_i^*/T]^{\kappa_i}} \left[\frac{p_{i^*i^*}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)}{\delta^* + p_{i^*i^*}^2(\mathbf{A}, \hat{\mathbf{\Omega}}_T)}\right]^{T/2} d\mathbf{A} \end{aligned}$$

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

Note next that

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

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

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

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

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

Given any $\varepsilon > 0$ and $\delta > 0$, by virtue of (62) 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 $\mathbf{M}_i = \mathbf{0}$, we have $\zeta_i^* = T\mathbf{a}_i'\hat{\mathbf{\Omega}}_T\mathbf{a}_i$ 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. Suggested standard priors for D and B.

Prior beliefs about structural variances should reflect in part the scale of the underlying data. Let $\hat{e}_{it}$ denote the residual of a mth-order univariate autoregression fit to 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 propose setting κ_i/τ_i (the prior mean for d_{ii}^{-1}) equal to the reciprocal of the ith diagonal element of $\mathbf{ASA}'$. Given equation (17), the prior is given a weight equivalent to $2\kappa_i$ observations of data; for example, setting $\kappa_i = 2$ (as is done in the empirical illustration in Section 5) would give the prior as much weight as 4 observations.

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. Let $\sqrt{s_{jj}}$ denote the estimated standard deviation of a univariate mth-order autoregression fit to variable j. Define

$$\underset{(1\times m)}{\mathbf{v}_1'} = \left(1/(1^{2\lambda_1}), 1/(2^{2\lambda_1}), ..., 1/(m^{2\lambda_1})\right) \tag{63}$$

$$\underset{(1\times n)}{\mathbf{v}_2'} = (s_{11}^{-1}, s_{22}^{-1}, ..., s_{nn}^{-1})'$$

$$\mathbf{v}_3 = \lambda_0^2 \begin{bmatrix} \mathbf{v}_1 \otimes \mathbf{v}_2 \\ \lambda_3^2 \end{bmatrix}. \tag{64}$$

Then $\mathbf{M}_i$ is taken to be a diagonal matrix whose row r column r element is the rth element of $\mathbf{v}_3$:

$$M_{i,rr} = v_{3r}. \tag{65}$$

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), and λ_3 is a separate parameter governing the tightness of the prior for the constant term, with all $\lambda_k \geq 0$.

Doan (2013) discussed possible values for these parameters. In the application in Section 5 we set $\lambda_1 = 1$ (which governs how quickly the prior for lagged coefficients tightens to zero as the lag ℓ increases), $\lambda_3 = 100$ (which makes the prior on the constant term essentially irrelevant), and set λ_0, the parameter controlling the overall tightness of the prior, to 0.2.

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 2 and verify directly the invariance claim used to produce equation (33).

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

To verify the invariance of the priors for impulse-response coefficients that are implicit in this algorithm in the absence of sign restrictions or other identifying assumptions, note from (28) that

$$\begin{aligned} h_{i1} &= p_{i1}q_{11}+p_{i2}q_{21}+\cdots+p_{ii}q_{i1} \\ &= \frac{p_{i1}x_{11}+p_{i2}x_{21}+\cdots+p_{ii}x_{i1}}{\sqrt{x_{11}^2+x_{21}^2+\cdots+x_{n1}^2}}. \end{aligned}$$

The numerator of h_{i1} is a zero-mean Normal variable with variance $p_{i1}^2+p_{i2}^2+\cdots+p_{ii}^2=\omega_{ii}$ and therefore the numerator can be written as $\sqrt{\omega_{ii}}v_{i1}$ where $v_{i1}\sim N(0,1)$ for

$$\begin{aligned} v_{i1} &= \frac{p_{i1}x_{11}+p_{i2}x_{21}+\cdots+p_{ii}x_{i1}}{\sqrt{\omega_{ii}}} \\ &= \boldsymbol{\alpha}_i'\mathbf{x}_1 \end{aligned}$$

for $\mathbf{x}_1$ the first column of $\mathbf{X}$ and $\boldsymbol{\alpha}_i'$ a vector with the property that $\boldsymbol{\alpha}_i'\boldsymbol{\alpha}_i=1$.

If we then consider the square of the impact,

$$h_{i1}^2 = \frac{\omega_{ii} v_{i1}^2}{x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2},$$

the claim is that this is ω_{ii} times a Beta(1/2,$(n-1)/2$) random variable. This would be the case provided that we can write the denominator as

$$x_{11}^2 + x_{21}^2 + \cdots + x_{n1}^2 = v_{11}^2 + v_{21}^2 + \cdots + v_{n1}^2$$

where $(v_{11}, ..., v_{n1})$ are independent $N(0,1)$ and the ith term v_{i1} is the same term as appears in the numerator. That this is indeed the case can be verified by noting that since $\mathbf{x}_1 \sim N(\mathbf{0}, \mathbf{I}_n)$, it is also the case that $\mathbf{v}_1 \sim N(\mathbf{0}, \mathbf{I}_n)$ when $\mathbf{v}_1 = \boldsymbol{\alpha}\mathbf{x}_1$ and $\boldsymbol{\alpha}$ is any orthogonal $(n \times n)$ matrix. We can start with any arbitrary unit-length vector $\boldsymbol{\alpha}_i'$ for the ith row of $\boldsymbol{\alpha}$ and always fill in the other rows $\{\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', ..., \boldsymbol{\alpha}_{i-1}', \boldsymbol{\alpha}_{i+1}', ..., \boldsymbol{\alpha}_n'\}$ so as to find such an orthogonal matrix $\boldsymbol{\alpha}$. Then $\mathbf{v}_1'\mathbf{v}_1 = \mathbf{x}_1'\boldsymbol{\alpha}'\boldsymbol{\alpha}\mathbf{x}_1 = \mathbf{x}_1'\mathbf{x}_1$ as claimed.

References

Akerlof, George A., and William T. Dickens (2007). "Unfinished Business in the Macroeconomics of Low Inflation: A Tribute to George and Bill by Bill and George," *Brookings Papers on Economic Activity* 2: 31-47.

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

Blanchard, Olivier J. and Peter Diamond (1990). "The Cyclical Behavior of Gross Flows of Workers in the United States," *Brookings Papers on Economic Activity* 1990(2): 85-155.

Blanchard, Olivier J., and Danny Quah (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* 79: 655-673.

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.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2013). "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macroeconomics Annual*, Volume 27, pp. 1-55, edited by Daron Acemoglu, Jonathan Parker, and Michael Woodford. Chicago, IL: University of Chicago Press.

Cho, Jang-Ok, and Thomas F. Cooley (1994). "Employment and Hours over the Business Cycle," *Journal of Economic Dynamics and Control* 18(2): 411-432.

Davis, Steven J. and John Haltiwanger (1999). "On the Driving Forces Behind Cyclical Movements in Employment and Job Reallocation," *American Economic Review* 89: 1234-1258.

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.

Devroye, Luc (1986). *Non-Uniform Random Variate Generation.* Springer-Verlag.

Drèze, Jacques H. (1974). "Bayesian Theory of Identification in Simultaneous Equations Models", in *Studies in Bayesian Econometrics and Statistics*, edited by S.E. Fienberg and A. Zellner. Amsterdam: North-Holland.

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

Faust, Jon and Eric M. Leeper (1997). "When Do Long-Run Identifying Restrictions Give Reliable Results?," *Journal of Business & Economic Statistics* 15: 345-353.

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

Galí, Jordi (1999). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?", *American Economic Review* 89: 249-271.

Galí, Jordi, Frank Smets, and Rafael Wouters (2012). "Unemployment in an Estimated New Keynesian Model," *NBER Macroeconomics Annual*, Volume 26, pp. 329-360, edited by Daron Acemoglu, and Michael Woodford. Chicago, IL: University of Chicago Press.

Giacomini, Raffaella, and Toru Kitagawa (2013). "Robust Bayes Inference in Non-Identified SVARs," working paper, University College London.

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

Gustafson, Paul (2009). "What Are the Limits of Posterior Distributions Arising From Nonidentified Models, and Why Should We Care?", *Journal of the American Statistical Asociation* 104: 1682-1695.

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

Hamermesh, Daniel S. (1996). *Labor Demand.* Princeton: Princeton University Press.

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

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.

Kydland, Finn E., and Edward C. Prescott (1982). "Time to Build and Aggregate Fluctuations," *Econometrica* 50(6): 1345-1370.

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

Leamer, Edward E. (1981). "Is it a Demand Curve, Or Is It A Supply Curve? Partial Identification through Inequality Constraints," *Review of Economics and Statistics* 63: 319-327.

Lichter, Andreas, Andreas Peichl, and Sebastian Siegloch (2014). "The Own-Wage Elasticity of Labor Demand: A Meta-Regression Analysis," working paper, Institute for the Study of Labor.

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

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

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.

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

Reichling, Felix, and Charles Whalen (2012). "Review of Estimates of the Frisch Elasticity of Labor Supply," working paper, Congressional Budget Office.

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.

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.

Tamer, Elie (2010). "Partial Identification in Econometrics," Annual Review of Economics 2: 167-195.

Theil, Henri (1971). *Principles of Econometrics.* New York: John Wiley & Sons, Inc.

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.

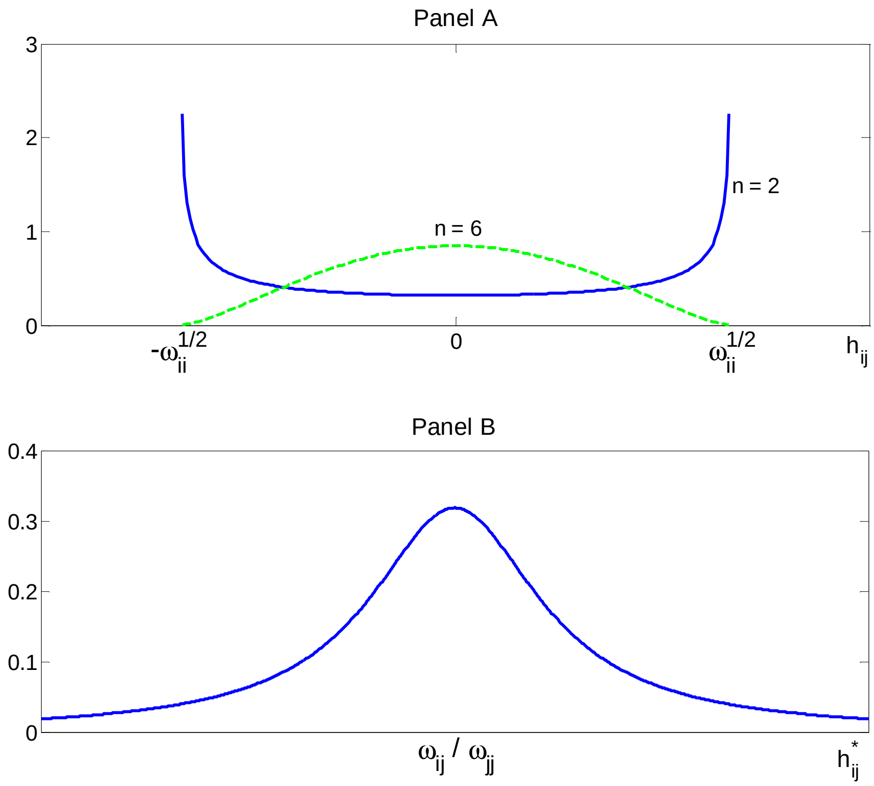

Figure 1. Prior densities for initial effect of shocks implicit in traditional approach to sign-restricted VAR. Panel A: response of variable *i* to one-standard-deviation increase of any structural shock when number of variables in VAR is 2 (solid blue) or 6 (dashed green). Panel B: response of variable *i* to a structural shock that increases variable *j* by one unit.

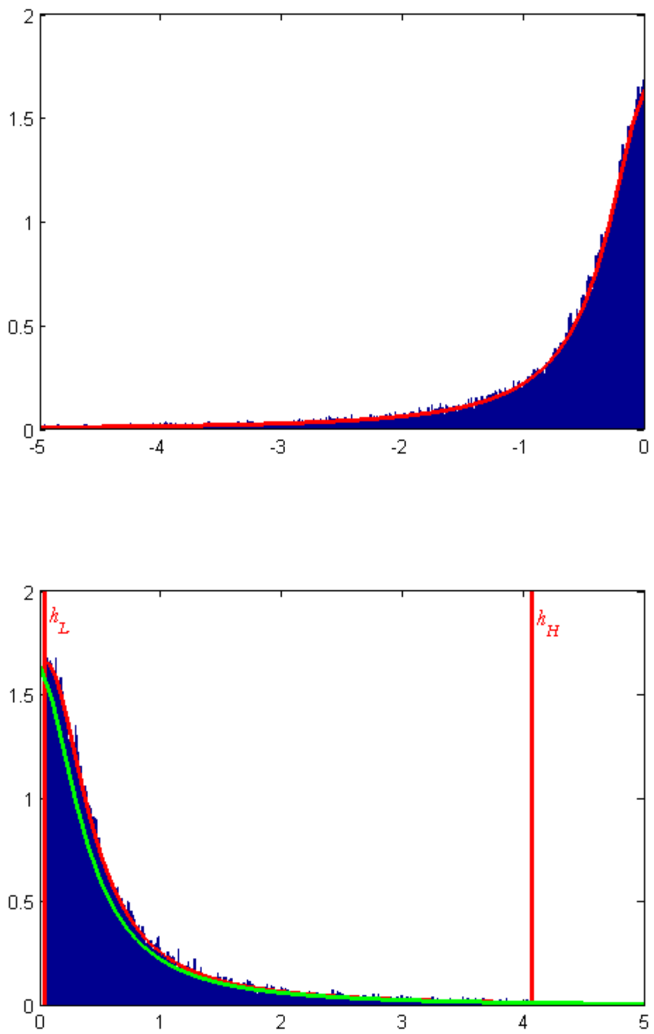

Figure 2. Effects of structural shocks as imputed by the traditional sign-restriction methodology (blue) and truncated priors implicit in that approach (red). Top panel: short-run wage elasticity of labor demand; bottom panel: short-run wage elasticity of labor supply.

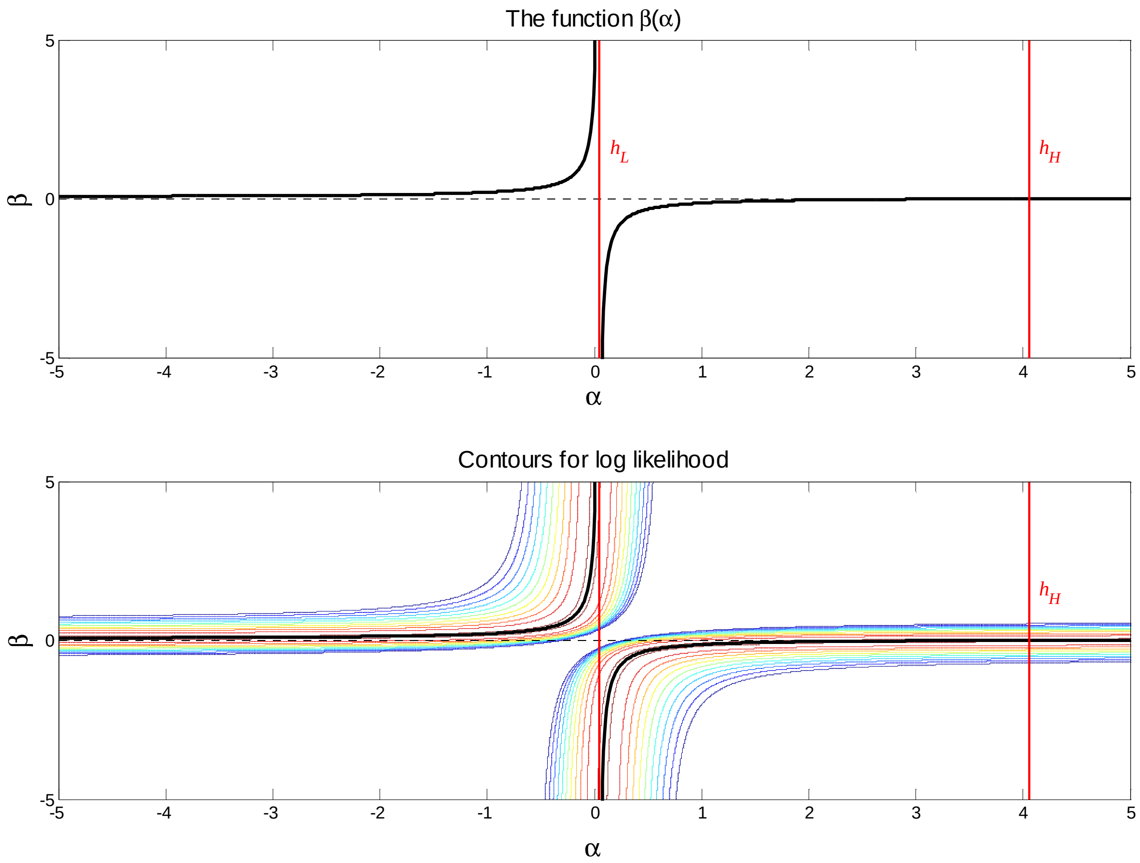

Figure 3. Maximum likelihood estimates and likelihood contours for α and β. Distance between contour lines is 10, and unshaded regions are exceedingly unlikely given the data.

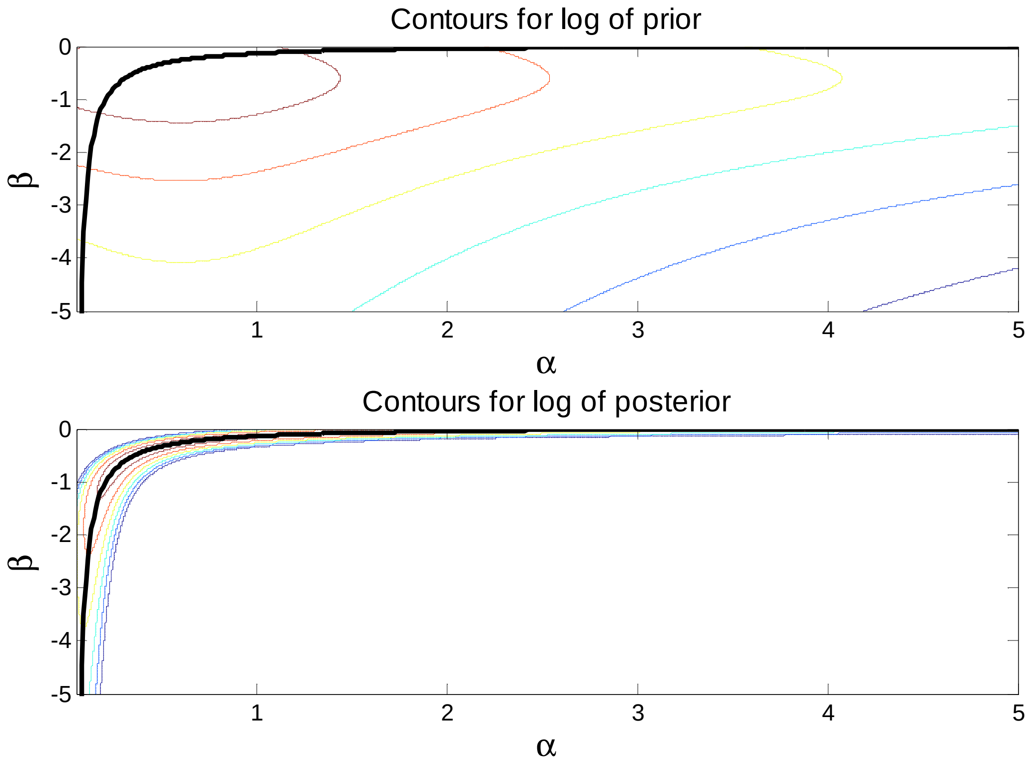

Figure 4. Maximum likelihood estimates and contours of prior and posterior distribution. Distance between contour lines is 2, and unshaded regions are exceedingly unlikely given prior beliefs and the data.

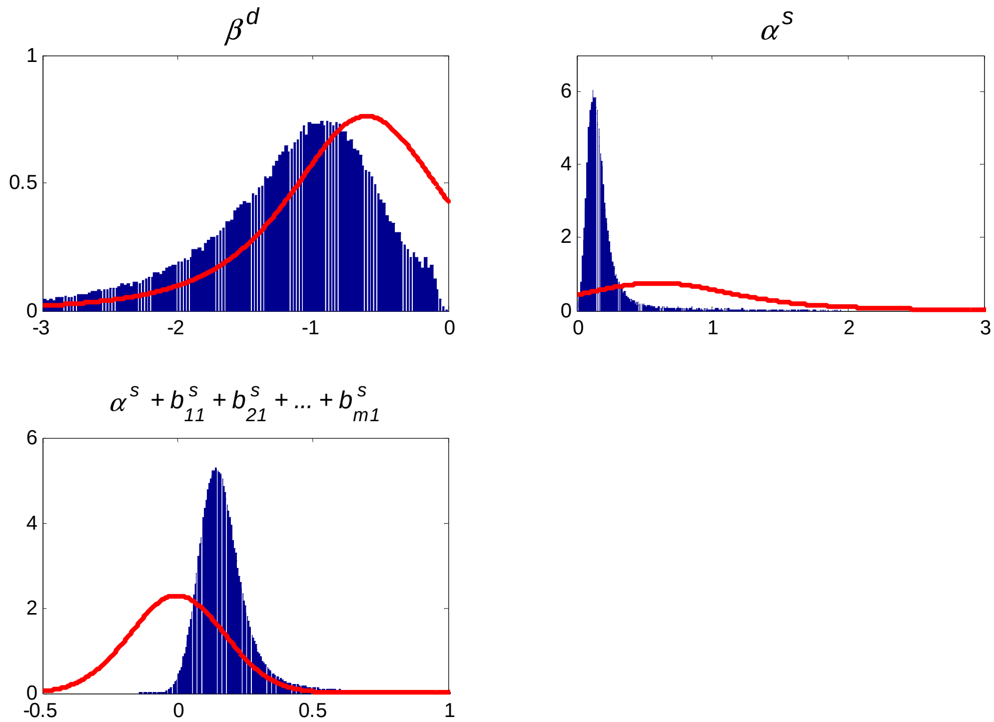

Figure 5. Distributions for prior (red curves) and posterior (blue histograms). Panel A: short-run elasticity of labor demand; panel B: short-run elasticity of labor supply; panel C: long-run labor-supply parameter (negative of lower left element of the matrix in equation (55)).

Figure 6. Posterior median (blue line) and 95% posterior credibility sets (shaded regions) for effects of labor demand and supply shocks on levels of employment and the real wage at alternative horizons (V = 0.1).

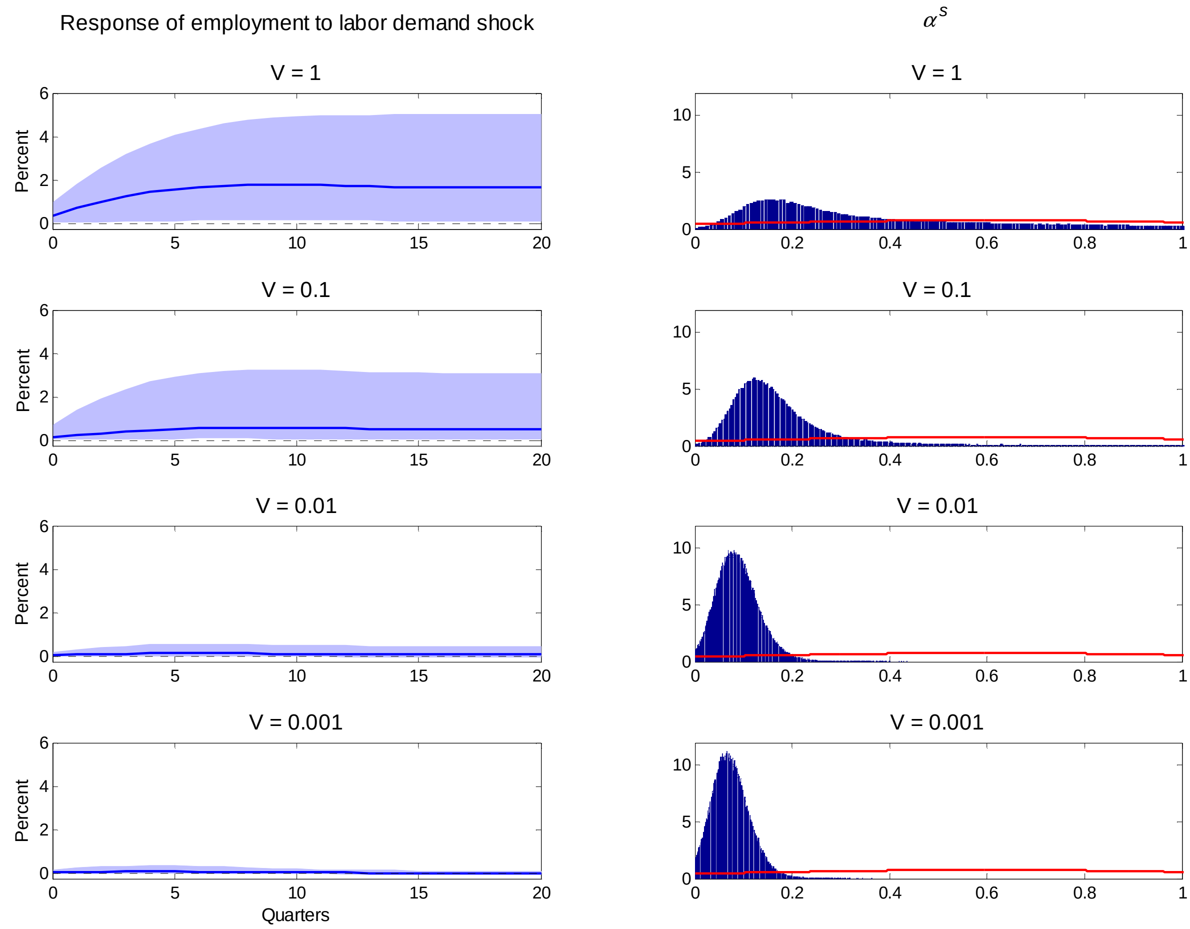

Figure 7. Effect of stronger priors about the long-run labor supply elasticity. Left column: response of employment to a labor demand shock for prior weights given by *V* = 1, 0.1, 0.01, or 0.001; right column: distribution for short-run labor supply elasticity for prior (red curve) and posterior (blue histograms) associated with different values of *V*.