Advances in Using Vector Autoregressions to Estimate Structural Magnitudes*

December 9, 2021
Revised: January 14, 2022

Christiane Baumeister
University of Notre Dame
cjsbaumeister@gmail.com

James D. Hamilton
University of California at San Diego
jhamilton@ucsd.edu

ABSTRACT

This paper discusses drawing structural conclusions from vector autoregressions. We call attention to a common error in estimating structural elasticities and show how to correctly estimate elasticities even in the case when one knows only the effects of a single structural shock and the covariance matrix of the reduced-form residuals. We describe the traditional approach to identification as a claim to have exact prior information about the structural model and propose Bayesian inference as a way to acknowledge that prior information is imperfect or subject to error. We raise concerns about the way that results are typically reported for VARs that are set-identified using sign and other restrictions.

JEL codes: C11, C32, Q43

Key words: structural vector autoregressions, Bayesian analysis, identification, elasticities, sign restrictions

*This paper supersedes earlier papers by the authors that were circulated under the titles "Advances in Structural Vector Autoregressions with Imperfect Identifying Information" and "Estimating Structural Parameters Using Vector Autoregressions." We are thankful to Frank Kleibergen, Lam Nguyen and Matthew Read for useful insights.

1 Introduction.

Vector autoregressions (VARs) offer a convenient tool for summarizing the dynamic correlations among a set of observed variables and are easily estimated by OLS regressions. In order to draw structural conclusions from those OLS regressions, we need to bring in some additional prior information about the economic structure. This paper reviews alternative approaches for doing this.

We distinguish between exact and inexact prior information. We define exact prior information as sufficient knowledge about the economic structure that would enable us to estimate structural magnitudes of interest with certainty if we had a large enough sample of data. Exact prior information is typically referred to as identifying assumptions. We illustrate inference with exact prior information using the familiar example of instrumental-variables estimation of the parameters of a demand equation. We show how maximum likelihood estimation of an exactly identified Gaussian structural VAR can be interpreted as an application of the principle of instrumental variables to obtain asymptotically optimal estimates of any structural magnitude of interest. We also comment on the role of parameterization, noting that maximum-likelihood estimates of any magnitude in a fully-identified structural VAR are invariant with respect to how the model is parameterized.

We use these results to call attention to a common error among applied researchers in trying to estimate a behavioral elasticity from the ratio of the estimated impacts of a single structural shock on different variables. We show that this procedure in general leads to inconsistent estimates even if the prior information is exact and correct. For models in which only the effects of a single structural shock are identified, we develop a formula that could be used to estimate consistently the parameters of that structural equation by combining knowledge of the effects of the structural shock with the observed covariance matrix of the reduced-form residuals.

We illustrate these ideas in a 3-variable VAR that is identified using a recursive structure (commonly called Cholesky identification). We note that if the demand equation is ordered last in the system, the identifying assumptions imply that the parameters of the demand equation can be estimated by an OLS regression of price on current quantity, income, and lagged values of the variables. We suggest that the implausibility of the resulting estimates raises doubts about the reliability of the identifying assumptions in that example.

We propose that the Bayesian approach to inference offers an appealing way to incorporate doubts that researchers may have about prior information and to acknowledge that the prior information is inexact. We review the algorithm developed by Baumeister and Hamilton (2015) for Bayesian analysis of structural VARs that takes advantage of natural conjugate distributions for a Gaussian structural VAR. We show that the traditional approach to identification can be viewed as a special case of Bayesian inference in which the researcher was certain before seeing the data that some of the parameters were zero but had no useful information about the remaining parameters. We show that a Bayesian approach can generalize the traditional approach to identification, using for illustration the case of a researcher who was extremely confident (but not completely certain) of the identifying information. We show how the inference changes as we move along the continuum from exact prior information to prior information in which the researcher has less confidence.

We propose that rather than claiming to have exact or extremely precise information about a few parameters, a better approach is to bring in inexact information from a variety of sources. We discuss how this can be done. We again note that the core issue is not how the model is parameterized but rather what aspects of the model we have prior information about.

We also discuss the relation between our recommended approach and structural vector autoregressions that are only set-identified on the basis of sign restrictions. We raise concerns about the common practice of users of sign-restricted VARs who highlight a single number as if it were the best estimate of some structural magnitude of interest and report error bands as if they summarized confidence in such estimates. We note that there is a Bayesian interpretation of the procedure that would justify this practice if the analysts' prior information about the structural model took a particular form, but note that published studies fail to articulate or defend the source of this prior information. We find a frequentist interpretation of the procedure to be even more problematic. The confidence bands used by practitioners are much too narrow from the perspective of a frequentist who is unpersuaded by the implicit prior information that underlies the popular methods.

Finally, we raise a caution about the algorithms used in sign-restricted VARs. In models where a large number of restrictions are imposed, it is sometimes the case that of the millions of draws generated only a handful are deemed suitable for inference. This raises questions about the efficiency and accuracy of the algorithm in such applications. We illustrate this using a prominent paper from the literature, demonstrating that if one simply changes the value of the seed from which the researchers' random numbers were generated, some of the key conclusions of the study would appear to be reversed.

2 Structural inference with exact identifying information.

In this section we discuss structural inference in the case when researchers have available identifying information that is exact in the sense that if we observed a large enough sample of data we could know structural parameters with certainty. We will illustrate some of the methods and issues using a simple example of the world oil market, taking one of the goals of the researcher being to estimate the price elasticity of the demand for crude oil.

2.1 Estimation by instrumental variables.

We follow every economics textbook in defining the price elasticity of demand as the response of buyers of the product to an increase in the price with other variables that influence demand held constant. Consider for example a dynamic demand equation in which q_t is a measure of the quantity of oil purchased, p_t is a measure of the real price of oil, and y_t is a measure of real income:

$$q_t = \delta y_t + \beta p_t + \mathbf{b}_d' \mathbf{x}_{t-1} + u_t^d. \tag{1}$$

In this equation, β is the short-run price elasticity of demand, δ is the short-run income elasticity of demand, u_t^d is a shock to demand, $\mathbf{x}_{t-1} = (1, \mathbf{y}_{t-1}', \mathbf{y}_{t-2}', ..., \mathbf{y}_{t-m}')'$ is a vector consisting of a constant term and m lags of each of the three variables with $\mathbf{y}_t = (q_t, y_t, p_t)'$, and $\mathbf{b}_d$ characterizes the response of demand to lagged values of the variables. The demand curve could be considered as part of a dynamic structural system that also describes the behavior of oil producers and the determinants of income:

$$q_t = \gamma y_t + \alpha p_t + \mathbf{b}_s' \mathbf{x}_{t-1} + u_t^s \tag{2}$$

$$y_t = \xi q_t + \psi p_t + \mathbf{b}_y' \mathbf{x}_{t-1} + u_t^y. \tag{3}$$

Here for example α is the short-run price elasticity of oil supply, u_t^s is a shock to oil production, and ψ is the contemporaneous effect of oil prices on economic activity.

We can write this structural model in vector form as

$$\mathbf{A}\mathbf{y}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u}_t \tag{4}$$

$$\mathbf{A} = \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 1 & -\delta & -\beta \end{bmatrix} \tag{5}$$

$$\mathbf{u}_t = (u_t^s, u_t^y, u_t^d)'$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_s' \\ \mathbf{b}_y' \\ \mathbf{b}_d' \end{bmatrix}.$$

As in most applications, we assume that the structural shocks have mean zero and are serially uncorrelated as well as uncorrelated with each other,

$$E(\mathbf{u}_t \mathbf{u}_s') = \begin{cases} \mathbf{D} & \text{for } t = s \\ \mathbf{0} & \text{for } t \neq s \end{cases},$$

with $\mathbf{D}$ diagonal.

Premultiplying (4) by $\mathbf{A}^{-1}$ results in the reduced-form VAR associated with the dynamic structural model:

$$\mathbf{y}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{6}$$

$$\mathbf{\Pi} = \mathbf{A}^{-1}\mathbf{B} \tag{7}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{A}^{-1}\mathbf{u}_t \tag{8}$$

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{A}^{-1}\mathbf{D}(\mathbf{A}^{-1})' = \mathbf{\Omega}. \tag{9}$$

Expression (8) establishes why we can't estimate the demand elasticities δ and β by OLS estimation of (1). From (8) and (6), when u_t^d goes up, it changes y_t and p_t. Thus although the demand shock u_t^d is uncorrelated with $\mathbf{x}_{t-1}$, it is correlated with the contemporaneous values of y_t and p_t. To estimate the parameters in (1) we need two instruments in addition to $\mathbf{x}_{t-1}$ that are correlated with y_t and p_t but uncorrelated with u_t^d.

Note that since the structural shocks are uncorrelated with each other, if we had consistent estimates of the other two structural shocks in the system $\hat{u}_t^s$ and $\hat{u}_t^y$, these could serve as valid instruments for estimating the demand equation, since they are correlated with y_t and p_t but uncorrelated with u_t^d.[1] IV estimates of the short-run demand elasticities could then be obtained from

$$\begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T \hat{u}_t^s \hat{\varepsilon}_t^y & \sum_{t=1}^T \hat{u}_t^s \hat{\varepsilon}_t^p \\ \sum_{t=1}^T \hat{u}_t^y \hat{\varepsilon}_t^y & \sum_{t=1}^T \hat{u}_t^y \hat{\varepsilon}_t^p \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T \hat{u}_t^s \hat{\varepsilon}_t^q \\ \sum_{t=1}^T \hat{u}_t^y \hat{\varepsilon}_t^q \end{bmatrix} \tag{10}$$

where $\hat{\boldsymbol{\varepsilon}}_t$ denote the residuals from OLS estimation of (6).

How could we get consistent estimates of the structural shocks $\hat{u}_t^s$ and $\hat{u}_t^y$? One answer would be if we had exact prior knowledge of the true values of the structural parameters $\gamma, \alpha, \xi,$ and ψ. In this case we could use

$$\begin{bmatrix} \hat{u}_t^s \\ \hat{u}_t^y \end{bmatrix} = \mathbf{\Gamma}\hat{\boldsymbol{\varepsilon}}_t$$

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix}.$$

This allows (10) to be expressed as

$$\mathbf{\Gamma}\hat{\mathbf{\Omega}}\hat{\boldsymbol{\eta}}_{IV} = \mathbf{0} \tag{11}$$

for $\boldsymbol{\eta} = (1, -\delta, -\beta)'$; for details, see Appendix A.

[1] By consistency here we mean that if $\hat{\theta}_T$ denotes an estimate of parameters based on a sample of size T, $\hat{u}_t^i(\hat{\theta}_T) - u_t^i \xrightarrow{p} 0$ and $T^{-1}\sum_{t=1}^T \hat{u}_t^i \hat{u}_t^j \xrightarrow{p} E(u_t^i u_t^j)$.

2.2 Estimation by full-information maximum likelihood.

If we assume that the structural shocks are distributed $N(\mathbf{0}, \mathbf{D})$, the log likelihood of the observed data $\mathbf{Y} = (\mathbf{y}_T', \mathbf{y}_{T-1}', ..., \mathbf{y}_1')'$ conditional on the pre-sample observations $\mathbf{x}_0$ is given by

$$\begin{aligned}\log f(\mathbf{Y}|\mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{x}_0) &= -(Tn/2)\log(2\pi) - (T/2)\log\left|\mathbf{A}^{-1}\mathbf{D}(\mathbf{A}^{-1})'\right| \\ &\quad - (1/2)\sum\nolimits_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}^{-1}\mathbf{B}\mathbf{x}_{t-1})'\left[\mathbf{A}^{-1}\mathbf{D}(\mathbf{A}^{-1})'\right]^{-1}(\mathbf{y}_t - \mathbf{A}^{-1}\mathbf{B}\mathbf{x}_{t-1}).\end{aligned} \tag{12}$$

If the VAR is stationary and the structural parameters $\mathbf{A}, \mathbf{B}, \mathbf{D}$ are identified, the values of $\hat{\mathbf{A}}_{MLE}, \hat{\mathbf{B}}_{MLE}, \hat{\mathbf{D}}_{MLE}$ that maximize (12) give the asymptotically optimal parameter estimates. Under these assumptions, the question of how to estimate the demand elasticity β has an unambiguous answer – we should use $\hat{\beta}_{MLE}$, which is given by the negative of the (3,3) element of $\hat{\mathbf{A}}_{MLE}$. If the observed sample $\mathbf{Y}$ and the identifying restrictions are the only information available, no consistent estimate of β can have a smaller asymptotic variance than $\hat{\beta}_{MLE}$.[2]

The likelihood could also be written in terms of the reduced-form parameters $\mathbf{\Pi}$ and $\mathbf{\Omega}$:

$$\begin{aligned}\log f(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Omega}, \mathbf{x}_0) &= -(Tn/2)\log(2\pi) - (T/2)\log|\mathbf{\Omega}| \\ &\quad - (1/2)\sum\nolimits_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Pi}\mathbf{x}_{t-1})'\mathbf{\Omega}^{-1}(\mathbf{y}_t - \mathbf{\Pi}\mathbf{x}_{t-1}).\end{aligned} \tag{13}$$

The maximum-likelihood estimates for this representation are given by

$$\hat{\mathbf{\Pi}}_{MLE} = \left(\sum\nolimits_{t=1}^{T}\mathbf{y}_t\mathbf{x}_{t-1}'\right)\left(\sum\nolimits_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1} \tag{14}$$

$$\hat{\mathbf{\Omega}}_{MLE} = T^{-1}\sum\nolimits_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t' \tag{15}$$

$$\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{\Pi}}_{MLE}\mathbf{x}_{t-1}.$$

The structural model is said to be just-identified if for any $\{\mathbf{\Pi}, \mathbf{\Omega}\}$ there exists a unique $\{\mathbf{A}, \mathbf{B}, \mathbf{D}\}$ satisfying (7), (9), and the structural restrictions imposed by the identifying assumptions. When the model is just-identified, the maximum-likelihood estimates of $\mathbf{A}$ and $\mathbf{D}$ are related to the MLE of $\mathbf{\Omega}$ as

$$\hat{\mathbf{D}}_{MLE} = \hat{\mathbf{A}}_{MLE}\hat{\mathbf{\Omega}}_{MLE}\hat{\mathbf{A}}_{MLE}'; \tag{16}$$

see Hamilton (1994, equation [11.6.33]). For the 3-equation example in Section 2.1 we can

[2] Under departures from the above assumptions – for example, in the presence of heteroskedasticity or structural breaks – other estimates could be much better than those obtained by maximization of (12). In that case the ideal procedure would be to model these features directly in the specification of the likelihood function and maximize the likelihood function that accurately characterizes the data.

write

$$\underset{(3\times 3)}{\mathbf{A}} = \begin{bmatrix} \underset{(2\times 3)}{\mathbf{\Gamma}} \\ \underset{(1\times 3)}{\boldsymbol{\eta}'} \end{bmatrix}.$$

Since $\mathbf{D}$ is diagonal, the (1,3) and (2,3) elements of (16) state that

$$\hat{\mathbf{\Gamma}}_{MLE}\hat{\mathbf{\Omega}}_{MLE}\hat{\boldsymbol{\eta}}_{MLE} = \mathbf{0}. \tag{17}$$

Comparing (17) with (11), it is clear that maximum likelihood estimation of the structural model (4) subject to the identifying restrictions is just a generalization of the familiar idea of estimation by instrumental variables.[3] MLE is the optimal application of the IV idea for the case when all the structural parameters are just-identified.

Other researchers prefer to use a representation in which structural shocks have unit variance, writing the system as

$$\mathbf{y}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \mathbf{H}\mathbf{u}_t^* \tag{18}$$

with $E(\mathbf{u}_t^*\mathbf{u}_t^{*\prime}) = \mathbf{I}_n$. In this representation, we are summarizing the structural shocks $\mathbf{u}_t^*$ in terms of how they influence the reduced-form residuals $\boldsymbol{\varepsilon}_t$:

$$\boldsymbol{\varepsilon}_t = \mathbf{H}\mathbf{u}_t^*.$$

The structural shocks $\mathbf{u}_t^*$ are interpreted to be identical to the structural shocks $\mathbf{u}_t$ in a system like (4) but scaled to have unit variance: $\mathbf{u}_t^* = \mathbf{D}^{-1/2}\mathbf{u}_t$. For example, $u_{t1}^*, u_{t2}^*, u_{t3}^*$ represent one-standard-deviation shocks to the supply, income, and demand equations respectively. The two representations are related by the identity

$$\mathbf{H} = \mathbf{A}^{-1}\mathbf{D}^{1/2}. \tag{19}$$

For the $\mathbf{H}$ representation, the log likelihood is

$$\begin{aligned} \log f(\mathbf{Y}|\mathbf{\Pi}, \mathbf{H}, \mathbf{x}_0) = &-(Tn/2)\log(2\pi) - (T/2)\log|\mathbf{H}\mathbf{H}'| \\ &- (1/2)\sum\nolimits_{t=1}^{T}(\mathbf{y}_t - \mathbf{\Pi}\mathbf{x}_{t-1})'(\mathbf{H}\mathbf{H}')^{-1}(\mathbf{y}_t - \mathbf{\Pi}\mathbf{x}_{t-1}). \end{aligned}$$

If the model is just-identified, the MLE of $\mathbf{H}$ is characterized by

$$\hat{\mathbf{H}}_{MLE}\hat{\mathbf{H}}'_{MLE} = \hat{\mathbf{\Omega}}_{MLE}. \tag{20}$$

Although this is a different parameterization, the concept of a demand elasticity as the

[3] To our knowledge, Shapiro and Watson (1988) were the first to point out the IV interpretation of maximum likelihood estimation of structural VARs.

response of buyers to an increase in price with income held constant is the same regardless of how we choose to represent the system. Recalling (19), if h^{ij} denotes the row i column j element of $\mathbf{H}^{-1}$, the maximum likelihood estimate of the short-run demand elasticity is

$$\hat{\beta}_{MLE} = -\hat{h}^{33}_{MLE}/\hat{h}^{31}_{MLE}. \tag{21}$$

For a just-identified model, this will be numerically identical to the estimate of this magnitude coming from the $(\mathbf{A}, \mathbf{B}, \mathbf{D})$ parameterization, which, as we noted above, is the optimal way to use observation of $\mathbf{Y}$ to estimate this magnitude.

2.3 A common error in estimating elasticities.

It has recently become a common practice among some applied researchers to report estimates that they interpret as supply or demand elasticities not on the basis of the inverse of $\mathbf{H}$ as called for in (21), but instead by calculating the ratios of the elements of a single column of $\mathbf{H}$ without inverting the matrix. For example, Kilian and Murphy (2014) proposed to estimate the elasticity of oil demand from the ratio of the change in oil consumption to the change in price in response to a shock to the supply of oil, that is, $\hat{\beta} = \hat{h}_{11}/\hat{h}_{31}$. Examples of other studies that have tried to estimate elasticities on the basis of the ratio of responses of different variables to a single structural shock include Kilian and Murphy (2012), Güntner (2014), Riggi and Venditti (2015), Kilian and Lütkepohl (2017), Ludvigson et al. (2017), Antolín-Díaz and Rubio-Ramírez (2018), Basher et al. (2018), Herrera and Rangaraju (2020), and Zhou (2020).

It is instructive to calculate the consequences of this procedure for the 3-equation example presented in Section 2.1. For that model, $\mathbf{H} = \mathbf{A}^{-1}\mathbf{D}^{1/2}$ with

$$\mathbf{A}^{-1} = |\mathbf{A}|^{-1} \begin{bmatrix} -\beta - \delta\psi & \alpha\delta - \beta\gamma & \alpha + \gamma\psi \\ -\psi - \beta\xi & \alpha - \beta & \psi + \alpha\xi \\ \delta\xi - 1 & \delta - \gamma & 1 - \gamma\xi \end{bmatrix}. \tag{22}$$

Thus

$$\frac{h_{11}}{h_{31}} = \frac{-\beta - \delta\psi}{\delta\xi - 1}. \tag{23}$$

In general, expression (23) is not the demand elasticity β. The reason is that if there is a one-standard-deviation shock to u^s_t, not only will it change the price p_t, but it will also change income. The size of the change in price is $\sqrt{d_{11}}|\mathbf{A}|^{-1}(\delta\xi - 1)$ and the size of the change in income is $\sqrt{d_{11}}|\mathbf{A}|^{-1}(-\psi - \beta\xi)$. From the demand curve, the change in price will lead to a change in quantity demanded of β times the change in price, namely $\beta\sqrt{d_{11}}|\mathbf{A}|^{-1}(\delta\xi - 1)$. Likewise the change in income will lead to a change in quantity demanded of δ times the change in income, namely $\delta\sqrt{d_{11}}|\mathbf{A}|^{-1}(-\psi - \beta\xi)$. The observed change in quantity demanded in response to the shock in supply is the sum of these two terms,

$$\underbrace{\beta}_{\text{response to price}} \underbrace{\sqrt{d_{11}}|\mathbf{A}|^{-1}(\delta\xi - 1)}_{\text{change in price}} + \underbrace{\delta}_{\text{response to income}} \underbrace{\sqrt{d_{11}}|\mathbf{A}|^{-1}(-\psi - \beta\xi)}_{\text{change in income}} = \underbrace{\sqrt{d_{11}}|\mathbf{A}|^{-1}(-\beta - \delta\psi)}_{\text{total change}}.$$

Dividing this by the magnitude of the change in price that results from the supply shock, $\sqrt{d_{11}}|\mathbf{A}|^{-1}(\delta\xi - 1)$, produces the result (23).

In the special case when demand does not respond to income ($\delta = 0$), expression (23) would simplify to the correct answer β. But in general, expression (23) reflects a combination of the sensitivity of demand to price, the sensitivity of demand to income, and the effects of an oil supply shock on those two variables.

Note that the correct calculation of elasticity that we gave in (21) is based on the ratios of elements in a row of the *inverse* of $\mathbf{H}$, not ratios of elements in a column of $\mathbf{H}$ itself.

Expression (23) does not summarize the characteristics of demand but instead characterizes the equilibrium impact of the structural shock. This is a fundamental problem for any study that attempts to calculate structural elasticities from the ratios of the effects of a particular structural shock. In a different example, Kilian and Murphy (2014) considered a 4-equation model in which there were two different kinds of demand shocks. These authors calculated the short-run price elasticity of supply in two different ways, first as the ratio of the change in quantity to the change in price resulting from the first demand shock, and second as the ratio of the change in quantity to the change in price resulting from the second demand shock. Kilian and Murphy supposed that either of these magnitudes could be regarded as estimates of the supply elasticity. In practice they will be two different numbers,[4] and neither corresponds to the usual understanding of what we mean by the supply elasticity, which is the parameter α in the structural equation (2).

2.4 Illustration: Cholesky identification.

We illustrate these results in a three-equation model of the world oil market of the form of (4) with $\mathbf{y}_t = (q_t, y_t, p_t)'$ for q_t the monthly growth rate of world oil production, y_t the monthly growth rate of world industrial production, and p_t the monthly growth rate of the dollar price of West Texas Intermediate crude oil deflated by the U.S. CPI.[5] We estimated the reduced-

[4] For example, running the code for Kilian and Murphy (2014) that is publicly posted in the *Journal of Applied Econometrics* data archive generates 5 million draws for the vector of possible parameters, of which the code discards all but 24,926 for reasons other than the supply elasticity. The code then calculates one supply elasticity from the ratio of the change in quantity to the change in price in response to what the authors interpret as an aggregate demand shock. The median value for this magnitude across the 24,926 draws is 0.1184. The code next calculates a second supply elasticity from the ratio of the change in quantity to the change in price in response to what the authors interpret as a speculative demand shock. The median value of this magnitude is 0.6184.

[5] The data are described in more detail in Baumeister and Hamilton (2019) and are posted at https://sites.google.com/site/cjsbaumeister/data_BH_AER2019.xlsx?attredirects=0&d=1.

form coefficients by OLS using (14) and (15) for $t = 1959$:M2 to 2019:M12, and calculated the reduced-form impulse response coefficients $\mathbf{\Psi}_s$ from the upper-left (3×3) block of $\mathbf{F}^s$ for $\mathbf{F}$ the companion matrix

$$\underset{(nm \times nm)}{\mathbf{F}} = \begin{bmatrix} \underset{(n \times nm)}{\tilde{\mathbf{\Pi}}} \\ \underset{[n(m-1) \times nm]}{\mathbf{E}} \end{bmatrix}$$

for $\tilde{\mathbf{\Pi}}$ the last nm columns of $\mathbf{\Pi}$ (dropping the intercepts in the equations) and $\mathbf{E} = \begin{bmatrix} \mathbf{I}_{n(m-1)} & \mathbf{0}_{[n(m-1) \times n]} \end{bmatrix}$.

For the structural model in this subsection we follow Kilian (2009) in assuming that the short-run income and price elasticities of supply (γ and α) as well as the contemporaneous coefficient relating oil prices to economic activity (ψ) are all zero. Under these structural assumptions, $\mathbf{A}$ and $\mathbf{H} = \mathbf{A}^{-1}\mathbf{D}^{1/2}$ are both lower triangular, and the MLE of $\mathbf{H}$ is given by the Cholesky factorization of $\hat{\mathbf{\Omega}}_{MLE}$:

$$\hat{\mathbf{H}}_{MLE} = \begin{bmatrix} 1.4334 & 0 & 0 \\ 0.0298 & 0.5458 & 0 \\ -0.2059 & 0.6366 & 6.9352 \end{bmatrix}. \tag{24}$$

The estimated effects on $\mathbf{y}_{t+s}$ of a one-standard-deviation-increase in one of the structural shocks at date t are then given by the matrix $\hat{\mathbf{\Psi}}_{s,MLE}\hat{\mathbf{H}}_{MLE}$.

We calculated standard errors for these estimates by generating draws of $\mathbf{\Omega}$ from an inverse-Wishart distribution with scale matrix $T\hat{\mathbf{\Omega}}_{MLE}$ and $T - mn - 1$ degrees of freedom. We then used this $\mathbf{\Omega}$ to draw vec($\mathbf{\Pi}'$) from a $N(\text{vec}(\hat{\mathbf{\Pi}}'), \mathbf{\Omega} \otimes [\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}'_{t-1}])$ distribution and also calculated the values of $\mathbf{\Psi}_s \times \text{chol}(\mathbf{\Omega})$ associated with this draw. This Normal-inverse-Wishart distribution can be motivated as the Bayesian posterior distribution if before seeing the data the Bayesian had an uninformative prior of the form proposed by Jeffreys (see Karlsson, 2013). It can alternatively be motivated as an approximation to the asymptotic frequentist distribution. Sims and Zha (1999) argued that this is also one of the best methods for approximating the small-sample frequentist distribution of impulse-response functions.

We plot estimates of the structural impulse-response functions along with 68% and 95% confidence bands in Figure 1. For comparability with earlier published studies, rows in the figure correspond to columns of $\mathbf{\Psi}_s\mathbf{H}$, and we report the effects of an oil-supply shock in terms of a one-standard-deviation *decrease* in oil production (-1 times the first column of $\mathbf{\Psi}_s\mathbf{H}$). The first row of Figure 1 shows that a decrease in oil production is followed by slower growth of world economic activity and a higher real price of oil. The (2,3) panel shows that higher economic activity is another factor that also leads to an increase in the real price of oil. Panel (3,3) shows that an increase in oil demand also leads to an increase in price.

We can also calculate the price elasticity of oil demand implied by this structural model.

Inverting the matrix in (24) and using expression (21), we find[6]

$$\hat{\beta}_{MLE} = -5.9562. \tag{25}$$

Again this estimate is invariant with respect to parameterization. Indeed, the easiest way to obtain the MLE is to divide equation (1) by β and write the result as[7]

$$\begin{aligned} p_t &= \beta^{-1} q_t - (\delta/\beta) y_t - \beta^{-1} \mathbf{b}_d' \mathbf{x}_{t-1} - \beta^{-1} u_t^d \\ &= \tilde{\beta} q_t + \tilde{\delta} y_t + \tilde{\mathbf{b}}_d' \mathbf{x}_{t-1} + \tilde{u}_t^d. \end{aligned} \tag{26}$$

Note that under the assumed recursive structural model, the error term in (26) is uncorrelated with any of the explanatory variables. The parameters can thus be immediately obtained by OLS, yielding $\tilde{\beta}_{MLE} = -0.167893$ which implies the identical estimate for $\hat{\beta} = 1/\tilde{\beta}$ as (25).

This estimated elasticity is astonishingly large. It implies that a 10% increase in the price of oil with no change in income would result in a 60% drop in consumption within the month. Monthly increases in oil prices of this magnitude are seen quite often, but monthly changes in consumption anywhere near 60% have never been observed. Baumeister and Hamilton (2019) noted that surveys of hundreds of studies using a wide range of methodologies systematically estimate the price elasticity of demand to be at least an order of magnitude smaller than this. Although a huge elasticity would be the conclusion we could draw from these data under these identifying assumptions, its implausibility might be one reason to doubt the validity of the identifying assumptions.

The assumption in the Cholesky identification that $\alpha = 0$ plays a key role in the conclusion that we could estimate the parameters of the demand equation by OLS estimation of (26). If instead it were the case that $\alpha > 0$, an increase in demand ($\tilde{u}_t^d > 0$) would result in an increase in price and thus an increase in the quantity of oil produced q_t. This would mean a positive correlation between the residual and the explanatory variable in (26). A positive correlation between between $\tilde{u}_t^d$ and q_t would bias the OLS estimate of $\tilde{\beta}$ in (26) upward; in other words, the plim of the OLS estimate would exceed the true $\tilde{\beta}$ by some positive number k. Since the true $\tilde{\beta}$ is negative, this makes the estimated $\tilde{\beta}$ closer to zero than it should be and thus biases the estimated demand elasticity $\hat{\beta} = \tilde{\beta}^{-1}$ toward a larger absolute value. Thus if the assumption in the Cholesky identification that $\alpha = 0$ is wrong, this could explain why when we impose this restriction that $\alpha = 0$ we get such an unreasonable estimate of the demand elasticity.

[6]If one instead tried to estimate the elasticity using the incorrect formula (23) the result would be -6.96.

[7]We could not estimate β from an OLS regression of the form of (1) with q_t on the left-hand side because u_t^d is correlated with p_t.

2.5 Inference when only a subset of parameters is identified.

Next we discuss inference when the researcher has exact prior information that would allow us to consistently estimate some but not all of the structural parameters. We illustrate this with a generalization of the example in the previous subsection in which we no longer impose a zero short-run response of supply to current income. However, in this subsection we will continue to assume zero short-run elasticities of supply and economic activity with respect to price. Thus the model we now consider is described by

$$\mathbf{A} = \begin{bmatrix} 1 & -\gamma & 0 \\ -\xi & 1 & 0 \\ 1 & -\delta & -\beta \end{bmatrix}. \tag{27}$$

The complete structural model is no longer identified, since there are 7 unknown structural parameters in (9), namely $\gamma, \xi, \delta, \beta, d_{11}, d_{22}, d_{33}$ but only 6 unique elements in the estimable symmetric matrix $\mathbf{\Omega}$. However, the elements of the third row of $\mathbf{A}$ can still be estimated. We can see this by noting that as long as the (1,3) and (2,3) elements of (27) are zero, the error in (26) is uncorrelated with q_t and y_t meaning we can still estimate (26) by OLS. The maximum likelihood estimates of β and δ are identical to those that were obtained in the fully identified case. The third column of $\mathbf{A}^{-1}$ still has two zeros in the first two rows, and the third column of $\hat{\mathbf{H}}$ will be identical to the third column of (24), with $\hat{h}_{33}$ the square root of the average squared residual from OLS estimation of (26) – the identical value as in the fully identified case.[8]

Thus consistent estimates of some parameters can still be obtained even if we relax the assumption that $\gamma = 0$, and in fact these estimates turn out to be numerically identical to those that resulted when we imposed $\gamma = 0$. However, the problem remains that if the short-run supply elasticity α is positive rather than zero, estimating the parameters of the demand equation by OLS results in an estimated short-run response of demand to the price that is biased in the direction of implying too large an absolute value for the elasticity.

2.6 Inference when only a single column of H is identified.

Is there a way to estimate the demand elasticity if we only have information about the effects of a single structural shock, and have no information at all, even imperfect or inexact information, about anything else in the system? This may arise for example in applications of the instrumental- or proxy-variable methods proposed by Stock and Watson (2012) and Mertens and Ravn (2013) when we only have available a proxy for one of the structural

[8] Another illustration of this result is that if we ordered the variables as (y_t, q_t, p_t) instead of the original (q_t, y_t, p_t) the third column of the Cholesky factor would be numerically identical. This point was originally made by Bernanke (1986) and Christiano, Eichenbaum, and Evans (1999).

shocks. The effects of the single shock are easy to estimate using the instrument in local projections (Plagborg-Møller and Wolf (2021)). Although we cannot infer structural elasticities from the jth column of $\mathbf{H} = \mathbf{A}^{-1}\mathbf{D}^{1/2}$ alone, it is possible to uncover the structural parameters of the jth row of $\mathbf{A}$ by using knowledge of the jth column of $\mathbf{H}$ together with $\boldsymbol{\Omega}$, the variance-covariance matrix of the reduced-form residuals. The jth column of $\mathbf{H}$ can be estimated (up to a constant of proportionality) as described in Stock and Watson (2012) and Mertens and Ravn (2013), and $\boldsymbol{\Omega}$ can be estimated from (15) without relying on any structural assumptions. This allows us to estimate parameters of the jth row of $\mathbf{A}$ under any chosen normalization for that equation, as we now demonstrate.[9]

It is easiest to derive this result for a normalization in which the structural shocks have unit variance, so that $\mathbf{A}^*\boldsymbol{\Omega}\mathbf{A}^{*\prime} = \mathbf{I}_n$. From this equation it follows that

$$\mathbf{A}^* = (\mathbf{A}^{*\prime})^{-1}\boldsymbol{\Omega}^{-1}. \tag{28}$$

Suppose we use proxy or instrumental variables to obtain an estimate $\hat{\mathbf{v}}_j$ of $\lambda_j \mathbf{h}_j$ where $\mathbf{h}_j$ denotes the effect of structural shock j (that is, $\mathbf{h}_j$ is the jth column of $\mathbf{A}^{*-1}$) and λ_j is an unknown constant of proportionality. The jth row of (28) allows us to estimate the jth row of $\mathbf{A}$ up to a constant of proportionality by

$$\hat{\mathbf{a}}_j^{*\prime} = \lambda_j^{-1}\hat{\mathbf{v}}_j^{\prime}\hat{\boldsymbol{\Omega}}^{-1}.$$

We can then express this in the $\mathbf{A}^*$ normalization by choosing λ_j so that $\hat{\mathbf{a}}_j^{*\prime}\hat{\boldsymbol{\Omega}}\hat{\mathbf{a}}_j^{*} = 1$, that is,[10]

$$\hat{\mathbf{a}}_j^{*\prime} = (\hat{\mathbf{v}}_j^{\prime}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{v}}_j)^{-1/2}\hat{\mathbf{v}}_j^{\prime}\hat{\boldsymbol{\Omega}}^{-1}. \tag{29}$$

To estimate the coefficients of the jth structural equation using a normalization such as expression (4), we would choose λ_j so that the appropriate element of $\hat{\mathbf{a}}_j$ was equal to unity. For example, to estimate the parameters δ and β in the demand curve (the third row of (5)), we would set λ_3 equal to the first element of $\hat{\mathbf{v}}_3^{\prime}\hat{\boldsymbol{\Omega}}^{-1}$ and estimate δ and β from the negative of the resulting second and third elements of $\lambda_3^{-1}\hat{\mathbf{v}}_3^{\prime}\hat{\boldsymbol{\Omega}}^{-1}$.

For the Cholesky example of Section 2.4, we would set $\hat{\mathbf{v}}_j$ equal to the third column of (24) and $\hat{\boldsymbol{\Omega}}$ the matrix in (15). For these values, expression (29) gives the estimates of parameters of the demand equation under the $\mathbf{A}^*$ normalization as

$$\hat{\mathbf{a}}_3^{*\prime} = \begin{bmatrix} 0.024209 & -0.168200 & 0.144193 \end{bmatrix}.$$

This is a third way to obtain the identical estimate $\hat{\beta} = -0.144193/0.024209 = -5.9562$ that

[9] We are in gratitude to Matthew Read for calling this point to our attention.

[10] Note that (29) satisfies $\hat{\mathbf{a}}_j^{*\prime}\hat{\boldsymbol{\Omega}}\hat{\mathbf{a}}_j^{*} = (\hat{\mathbf{v}}_j^{\prime}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{v}}_j)^{-1/2}\hat{\mathbf{v}}_j^{\prime}\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{v}}_j(\hat{\mathbf{v}}_j^{\prime}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{v}}_j)^{-1/2} = 1$.

we earlier arrived at using either expression (21) or (26).

Note that our recommended approach for using the effects of a single structural shock to estimate parameters of a structural equation differs in two fundamental respects from the approach we criticized in Section 2.3. First, our expression (29) makes use of the covariance matrix of the reduced-form residuals $\hat{\mathbf{\Omega}}$ in addition to a single column of $\mathbf{H}$, whereas an expression like (23) claims to be able to find the answer from $\mathbf{h}_j$ alone. Second, our expression uses knowledge of the effects of a demand shock to estimate the parameters of the demand equation, whereas (23) claims to be able to estimate the parameters of the demand equation based on the effects of a supply shock alone.

3 Structural inference with inexact identifying information.

In the previous section we took the view that prior to seeing the data, the analyst had exact prior information about some aspects of the structural model. For example, in Section 2.5 we assumed we knew with certainty that $\alpha = \psi = 0$. Bayesian analysis offers a natural way to incorporate doubts about the reliability of this kind of prior information into the conclusions we draw from the data.

3.1 Calculating the Bayesian posterior distribution in structural vector autoregressions.

Let $\boldsymbol{\psi}$ be a vector of parameters containing the unknown elements of $(\mathbf{A}, \mathbf{D}, \mathbf{B})$. Whereas a frequentist would summarize prior information about $\boldsymbol{\psi}$ in the form of some restrictions on the parameters that we know with certainty must hold, the Bayesian represents prior information in the form of a probability density $p(\boldsymbol{\psi})$. This density is higher for values of $\boldsymbol{\psi}$ that economic theory or analysis of previous data sets led us to think are more plausible, and lower for values that are less consistent with earlier findings. The goal of the researcher is to use the distribution of the data given those parameters (that is, the likelihood $f(\mathbf{Y}|\boldsymbol{\psi})$) along with Bayes' Law to evaluate the plausibility of different parameters after having seen the data. This is summarized in terms of a posterior density $p(\boldsymbol{\psi}|\mathbf{Y})$:

$$p(\boldsymbol{\psi}|\mathbf{Y}) = \frac{p(\boldsymbol{\psi}) f(\mathbf{Y}|\boldsymbol{\psi})}{\int p(\boldsymbol{\psi}) f(\mathbf{Y}|\boldsymbol{\psi}) d\boldsymbol{\psi}}.$$

Numerical advances in principle allow us to calculate $p(\boldsymbol{\psi}|\mathbf{Y})$ for an arbitrary prior $p(\boldsymbol{\psi})$ and likelihood function $f(\mathbf{Y}|\boldsymbol{\psi})$. But given the large numbers of parameters in VARs, it is helpful to use distributions that allow us to perform most of the calculation analytically.

Baumeister and Hamilton (2015) suggested one way to do this. The ith row of (4) states

$$\mathbf{a}_i'\mathbf{y}_t = \mathbf{b}_i'\mathbf{x}_{t-1} + u_{it} \tag{30}$$

with $u_{it} \sim N(0, d_{ii})$ independent across i and t. If we knew the value of $\mathbf{A}$, this would be a standard Gaussian regression model of $\mathbf{a}_i'\mathbf{y}_t$ on $\mathbf{x}_{t-1}$ for which analytical results for Bayesian inference about d_{ii} and $\mathbf{b}_i$ are well known. Specifically, given $\mathbf{A}$, if the prior for $(d_{ii},\mathbf{b}_i)$ was described by a Normal-inverse-gamma distribution independent across i, then the posterior distribution of $(d_{ii},\mathbf{b}_i)$ conditional on $\mathbf{A}$ is in the same class of distributions and known analytically. This allows us to concentrate the numerical part of the inference on $\mathbf{A}$ alone, which has at most $n^2 - n$ unknown elements.

Baumeister and Hamilton's approach allows the prior distribution for the contemporaneous structural coefficients $p(\mathbf{A})$ to be represented by any proper density that can be evaluated numerically up to a possibly unknown multiplicative constant. Values of $\mathbf{A}$ that prior evidence or theory suggests are more plausible are associated with a bigger value for $p(\mathbf{A})$ and values that can be ruled out altogether are represented by $p(\mathbf{A}) = 0$.

Baumeister and Hamilton suggested using an inverse-gamma distribution for the prior on diagonal elements of $\mathbf{D}$ conditional on $\mathbf{A}$:

$$p(d_{ii}^{-1}|\mathbf{A}) = \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}(d_{ii}^{-1})^{\kappa_i-1}\exp(-\tau_i d_{ii}^{-1}) \quad \text{for } d_{ii}^{-1} \geq 0.$$

This turns out to be the natural-conjugate prior in the sense that if the prior is of this form, then the posterior distribution turns out also to be of this form. The confidence in this prior is governed by the parameter κ_i. In the illustrations in this paper, we use $\kappa_i = 0.5$ which weights the prior as equivalent to the information that would come from a single observation on $\mathbf{y}_t$. The value of τ_i/κ_i represents the value of d_{ii} that we might have anticipated before seeing the data. We follow Doan, Litterman, and Sims (1984) in basing this on the average squared residuals from univariate autoregressions for the individual elements of $\mathbf{y}_t$.[11]

Baumeister and Hamilton used a prior for the ith row of $\mathbf{B}$ conditional on $\mathbf{A}$ and $\mathbf{D}$ that is Normally distributed with mean $\mathbf{m}_i(\mathbf{A})$ and variance $d_{ii}\mathbf{M}_i$. Thus $\mathbf{m}_i(\mathbf{A})$ summarizes the value of $\mathbf{b}_i$ that we expected before seeing the data and $\mathbf{M}_i$ summarizes our confidence in this guess, with smaller diagonal elements of $\mathbf{M}_i$ corresponding to more confidence in the prior information about that parameter. This again is the natural-conjugate prior. We use the "Minnesota prior" in Doan, Litterman, and Sims (1984) and Sims and Zha (1998) to characterize the mean $\mathbf{m}_i(\mathbf{A})$ and variance $d_{ii}\mathbf{M}_i$ of this distribution. In the applications in this paper we set the parameter that governs the overall scale of $\mathbf{M}_i$ at $\lambda_0 = 10^9$, which

[11] Specifically, let $\hat{v}_{it}$ be the residuals from an OLS regression of y_{it} on $(1, y_{i,t-1}, ..., y_{i,t-m})'$. Collect these in an $(n \times 1)$ vector $\hat{\mathbf{v}}_t$ and calculate $\hat{\mathbf{S}} = T^{-1}\sum_{t=1}^{T}\hat{\mathbf{v}}_t\hat{\mathbf{v}}_t'$. As in (16) we set $\bar{d}_{ii}$ to be the ith diagonal element of $\mathbf{A}\hat{\mathbf{S}}\mathbf{A}'$ and $\tau_i = \kappa_i\bar{d}_{ii}$.

represents an essentially uninformative prior distribution for $\mathbf{B}$. The complete prior is then $p(\mathbf{A}, \mathbf{D}, \mathbf{B}) = p(\mathbf{A})p(\mathbf{D}|\mathbf{A})p(\mathbf{B}|\mathbf{A}, \mathbf{D})$.

Conditional on $\mathbf{A}$, the posterior distributions of $\mathbf{D}$ and $\mathbf{B}$ can be calculated using well known formulas. Given a value for $\mathbf{A}$, define

$$s_i^{\mathbf{yy}}(\mathbf{A}) = \mathbf{a}_i' \textstyle\sum_{t=1}^T \mathbf{y}_t \mathbf{y}_t' \mathbf{a}_i + \mathbf{m}_i(\mathbf{A})' \mathbf{M}_i^{-1} \mathbf{m}_i(\mathbf{A})$$

$$s_i^{\mathbf{yx}}(\mathbf{A}) = \mathbf{a}_i' \textstyle\sum_{t=1}^T \mathbf{y}_t \mathbf{x}_{t-1}' + \mathbf{m}_i(\mathbf{A})' \mathbf{M}_i^{-1}$$

$$s_i^{\mathbf{xx}} = \textstyle\sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}_{t-1}' + \mathbf{M}_i^{-1}$$

$$\zeta_i^*(\mathbf{A}) = s_i^{\mathbf{yy}}(\mathbf{A}) - s_i^{\mathbf{yx}}(\mathbf{A})(s_i^{\mathbf{xx}})^{-1} s_i^{\mathbf{yx}}(\mathbf{A})'$$

$$\tau_i^*(\mathbf{A}) = \tau_i(\mathbf{A}) + (1/2)\zeta_i^*(\mathbf{A})$$

$$\mathbf{m}_i^*(\mathbf{A}) = (s_i^{\mathbf{xx}})^{-1} s_i^{\mathbf{yx}}(\mathbf{A})'$$

$$\mathbf{M}_i^*(\mathbf{A}) = (s_i^{\mathbf{xx}})^{-1}.$$

These expressions could be calculated if desired using a standard regression package. For example, $\mathbf{m}_i^*(\mathbf{A})$ can be written as $(\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)^{-1} \tilde{\mathbf{X}}_i' \tilde{\mathbf{y}}_i$ where $\tilde{\mathbf{X}}_i$ is a $(T + k) \times k$ matrix whose first T rows contain the values of $\mathbf{x}_{t-1}'$ and whose last k rows are given by $\mathbf{P}_i$ which denotes the Cholesky factor of $\mathbf{M}_i^{-1}$. Likewise $\tilde{\mathbf{y}}_i$ is a $(T + k) \times 1$ vector whose first T elements are the values of $\mathbf{a}_i' \mathbf{y}_t$ and whose last k elements are $\mathbf{m}_i(\mathbf{A})$.

Using these expressions, Baumeister and Hamilton showed that the posterior distribution of $\mathbf{A}$ conditional on the data $\mathbf{Y}$ and the presample observations $\mathbf{x}_0$ is

$$p(\mathbf{A}|\mathbf{Y}, \mathbf{x}_0) \propto \frac{p(\mathbf{A})[\det(\mathbf{A}\hat{\mathbf{\Omega}}_{MLE}\mathbf{A}')]^{T/2}}{\prod_{i=1}^n [(2/T)\tau_i^*(\mathbf{A})]^{\kappa_i^*}} \prod_{i=1}^n \tau_i(\mathbf{A})^{\kappa_i}. \tag{31}$$

This can be calculated for any given $\mathbf{A}$, allowing us to use numerical Bayesian methods to find the posterior Bayesian distribution of $\mathbf{A}$. Since this has at most $n^2 - n$ unknown elements, this greatly reduces the dimensionality of the problem. Baumeister and Hamilton used a random-walk Metropolis Hastings algorithm to generate draws of the unknown elements of $\mathbf{A}$ from the distribution in (31).

The posterior distribution of diagonal elements of $\mathbf{D}$ conditional on $\mathbf{A}$ turns out to be inverse-gamma with parameters $\kappa_i^* = \kappa_i + T/2$ and $\tau_i^*(\mathbf{A})$ given in the expression above. Thus given a draw for $\mathbf{A}$ from the distribution in (31) we can generate a draw from $p(\mathbf{D}|\mathbf{A}, \mathbf{Y}, \mathbf{x}_0)$ analytically. The posterior distribution for the ith row of $\mathbf{B}$ conditional on $\mathbf{A}$ and $\mathbf{D}$ turns out to be $N(\mathbf{m}_i^*(\mathbf{A}), d_{ii}\mathbf{M}_i^*)$, which can again be done analytically. Baumeister and Hamilton's algorithm to generate draws from the posterior distribution of parameters $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y})$ and the complete structural impulse-response function $p\left(\{\mathbf{\Psi}_s \mathbf{A}^{-1}\}_{s=0}^{\infty} \middle| \mathbf{Y}\right)$ is available at https://drive.google.com/uc?export=download&id=1Dh34goEM1neTjykrwdpa60jvyU92nBg1.

3.2 Optimal Bayesian estimates.

Statistical decision theory can be used to summarize the posterior distributions in terms of point estimates of any magnitudes of interest. The goal of the analyst is taken to be to choose as an estimate of some magnitude of interest $\boldsymbol{\theta}$ the value that minimizes expected posterior loss,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\mathbf{g}} \int \ell(\boldsymbol{\theta}, \boldsymbol{g}) p(\boldsymbol{\theta}|\mathbf{Y}) d\boldsymbol{\theta},$$

where $\ell(\boldsymbol{\theta}, \boldsymbol{g})$ denotes the cost of relying on $\mathbf{g}$ as an estimate when the true value is $\boldsymbol{\theta}$**.** For example, Baumeister and Hamilton (2018) demonstrated that if the analyst has a quadratic loss function defined over the full impulse-response function, $\ell(\boldsymbol{\theta}, \boldsymbol{g}) = (\boldsymbol{\theta} - \boldsymbol{g})'\mathbf{W}(\boldsymbol{\theta} - \boldsymbol{g})$ for $\boldsymbol{\theta} = \text{vec}\left(\begin{bmatrix} \boldsymbol{\Psi}_0\mathbf{A}^{-1} & \boldsymbol{\Psi}_1\mathbf{A}^{-1} & \cdots & \boldsymbol{\Psi}_h\mathbf{A}^{-1} \end{bmatrix}\right)$ and $\mathbf{W}$ any positive semidefinite matrix, the optimal estimate of the ith element of $\boldsymbol{\theta}$ is obtained from the point-by-point average value of θ_i across draws from $p(\boldsymbol{\theta}|\mathbf{Y})$. Alternatively, if the loss function is any weighted sum of the absolute values of individual elements $(\theta_i - g_i)$ with nonnegative weights, the optimal posterior estimate of the impulse-response function is obtained by calculating the point-by-point median values of draws from the posterior distribution of the structural impulse-response function.

Fry and Pagan (2011) suggested that when a researcher reports an estimate of an impulse-response function, it is desirable that every element in the reported function should result from the same value for the triple $(\mathbf{A}, \mathbf{D}, \mathbf{B})$. Inoue and Kilian (forthcoming) tried to give this claim a decision-theoretic foundation by introducing a loss function that takes on the value of infinity unless every element of $\boldsymbol{\theta}$ comes from a single value $(\mathbf{A}, \mathbf{D}, \mathbf{B})$. It is far from clear what decision a user of research would ever have to make that would involve a payoff function of this form. The elements of $\boldsymbol{\theta}$ are different functions of $(\mathbf{A}, \mathbf{D}, \mathbf{B})$ and there will be different levels of posterior uncertainty associated with those different elements. In general, the optimal estimate of α^2 is not the square of the optimal estimate of α. For most loss functions, the optimal estimate $\mathbf{g}$ of the vector $\boldsymbol{\theta} = (\alpha, \alpha^2)'$ will not have the property that the second element of $\mathbf{g}$ is the square of the first. Insofar as the goal of Fry and Pagan's criterion is to achieve internal consistency, this the standard Bayesian approach accomplishes by construction. The Bayesian describes the data with a single, internally consistent model, namely the likelihood function $f(\mathbf{Y}|\mathbf{A}, \mathbf{D}, \mathbf{B})$, and summarizes prior information about $\mathbf{A}, \mathbf{D}, \mathbf{B}$ using a single, internally consistent density in the form of the prior $p(\mathbf{A}, \mathbf{D}, \mathbf{B})$. Bayes' Law allows us to characterize our uncertainty after seeing the data in the form of a unique and well defined posterior distribution $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y})$. With this posterior distribution we can characterize our posterior uncertainty about any function of $\mathbf{A}, \mathbf{D}, \mathbf{B}$ such as $\boldsymbol{\theta}(\mathbf{A}, \mathbf{D}, \mathbf{B})$. This fully internally consistent posterior distribution is exactly what is generated by Baumeister and Hamilton's algorithm.

3.3 A Bayesian interpretation of the traditional approach to identification.

A large literature has explored the conditions under which Bayesian inference can encompass the methods used by frequentists as a special case; see for example Phillips (1991), Chao and Phillips (2002), and Kleibergen and Zivot (2003). These treatments assume that the Bayesian accepts the same identifying restrictions used by the frequentist and in addition brings in information about other structural parameters in the form of a Bayesian prior distribution. A typical result is that encompassing obtains if the prior distribution takes the form suggested by Jeffreys (1946). The numerical algorithm described in Section 3.1 does not include the Jeffreys prior as a special case. Our algorithm assumes that the prior distribution is a proper density that integrates to unity, which the Jeffreys prior does not.

Here we discuss a different sense in which the frequentist approach can be described as a special case of Bayesian inference. The identifying assumptions used by the frequentist can be viewed as making use of prior information about the structural model, namely, features of the structural model that the analyst claims to know with certainty before seeing the data. A generalization of this approach is that these restrictions are very likely true, but we are not completely certain. This Bayesian view of identification was proposed by Lindley (1971) and Drèze (1975), but to our knowledge did not see much use by applied econometricians until the work of Baumeister and Hamilton (2015). In this section we follow Baumeister and Hamilton (2019) in showing how we can interpret the traditional approach to identification as a special case of Bayesian inference.

The Cholesky identification in Section 2.4 could be described as the inference of a Bayesian who had no useful prior information about the (2,1), (3,1), and (3,2) elements of $\mathbf{A}$, but was certain before seeing the data (and therefore will remain certain after seeing the data) that the (1,2), (1,3), and (2,3) elements of $\mathbf{A}$ are all zero. We implemented this using the Bayesian algorithm described in Section 3.1 in which the prior distribution allowed zero possibility that the (1,2), (1,3), or (2,3) values could be nonzero. The prior for the (2,1), (3,2) and (3,3) elements of $\mathbf{A}$ was represented by independent Student t distributions with location parameter 0, scale parameter 100, and 3 degrees of freedom. This prior for the unknown elements of $\mathbf{A}$ is essentially flat over any conceivable range for these parameters. We also use relatively uninformative priors for $p(\mathbf{D}|\mathbf{A})$ and $p(\mathbf{B}|\mathbf{A},\mathbf{D})$ as described in Section 3.1.

We summarize the posterior distribution of the response of p_{t+s} to a one-standard-deviation shock to u_t^s, u_t^y, or u_t^d in the first column of Figure 2. The horizontal axis plots the horizon s and the height of the solid blue line is the median of the posterior distribution of the response at that horizon. Bands display 68% and 95% credibility sets of the posterior distribution. Not surprisingly, these are identical to the results in the third column of Figure 1, which represented inference of a frequentist econometrician who relied on Cholesky identification.

For ease of comparison, we plot the Cholesky maximum likelihood estimates from Figure 1 as dotted red lines in Figure 2. These coincide exactly with the Bayesian posterior medians in solid blue. Thus we could describe the traditional approach to identification as a special case of Bayesian inference in which the analyst has exact prior knowledge about some elements of the structure, in this case, exact knowledge that γ, α, and ψ are all zero, but no useful prior knowledge about any other elements of the structure.

The first panel of Figure 3 plots the posterior distribution of the demand elasticity β that results from Bayesian inference using this prior. This distribution assigns a 17.5% probability to a positive value (i.e., to the claim that an increase in price leads to an *increase* in the quantity demanded) and a 97% probability to a value greater than two in absolute value (i.e., a 10% increase in price leads to more than a 20% change in quantity demanded). These outcomes are highly implausible, but are allowed because the Bayesian prior used in this subsection makes no use of any prior information about β. This highlights a striking asymmetry in the Cholesky approach to identification. Cholesky identification claims that we know with certainty the value of the supply elasticity (which we surely don't) and yet know nothing at all about the value of the demand elasticity (which we surely do).

3.4 A Bayesian generalization of the traditional approach to identification.

We now discuss Bayesian inference in the case when we have some doubts about the validity of the identifying assumptions. We use as an example an analyst who may not be completely certain that the short-run supply elasticity α is exactly zero. Consider the following generalization of the Cholesky specification:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -\alpha \\ -\xi & 1 & 0 \\ 1 & -\delta & -\beta \end{bmatrix}. \tag{32}$$

We assume as in the previous example that the econometrician has no useful prior information about the value of ξ or the demand parameters δ and β. In contrast to the previous example, the analyst now has inexact but potentially still highly informative prior information about the value of α. We assume that the analyst knows with certainty that the supply elasticity cannot be negative nor greater than some upper bound α_0, with the prior distribution uniform over α within these bounds. Thus the prior density for this example is

$$p(\mathbf{A}) \propto \begin{cases} \left[1 + \frac{1}{\nu}\left(\frac{\xi}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} \left[1 + \frac{1}{\nu}\left(\frac{\delta}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} \left[1 + \frac{1}{\nu}\left(\frac{\beta}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} & \text{if } \alpha \in [0, \alpha_0] \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

with $\nu = 3$ and $\sigma = 100$.

Note that this is a strict generalization of Cholesky identification, getting arbitrarily close to the traditional recursive model as $\alpha_0 \to 0$. For example, Kilian and Murphy (2012) argued that the supply elasticity cannot be any larger than 0.025, meaning that a 10% increase in price would lead to an increase in production that is less than 0.25%. The second column of Figure 2 summarizes the Bayesian posterior distribution of the impulse-response functions when the prior distribution is given by (33) with $\alpha_0 = 0.025$. These are very similar to the full Cholesky case in column 1 because the upper bound α_0 is so close to zero. However, a researcher who allowed some possibility of a very small positive supply elasticity would assign a higher posterior probability to bigger price effects of oil supply disruptions than the analyst who was certain that $\alpha = 0$.

Although the graphs in the first two columns of Figure 2 appear similar to each other, there is one important conceptual distinction. The error bands in the first panel reflect only estimation uncertainty about the reduced-form VAR parameters $\mathbf{\Pi}$ and $\mathbf{\Omega}$. As the sample size T goes to infinity, we would be able to estimate these parameters without error. Since the mapping from $\mathbf{\Pi}$ and $\mathbf{\Omega}$ to the structural parameters $(\mathbf{A}, \mathbf{D}, \mathbf{B})$ is also known with certainty for the specification in the first column, the error bands in the first column would collapse to the point estimates as the sample size becomes infinite. In the second column, by contrast, even if the sample size T were infinite and we knew the values of $\mathbf{\Pi}$ and $\mathbf{\Omega}$ with certainty, we would still not have complete confidence in our knowledge about structural magnitudes like the effects of a supply disruption because we have some uncertainty about the identification itself. Unlike the first column of Figure 2, the error bands in the second column of Figure 2 reflect both estimation uncertainty and uncertainty about the identification. The latter uncertainty would remain even if the sample size were infinite.

The second panel of Figure 3 displays the posterior distribution of the demand elasticity that results from this less dogmatic prior on the supply elasticity. Interestingly, although the prior for the demand elasticity is just as uninformative as in the first case, relaxing the dogmatic prior on the supply elasticity results in a significantly more reasonable posterior distribution for the demand elasticity. The posterior probability that the demand elasticity could be positive is now only 2.9% (not visible given the scale of the figure), and the probability of a value greater than two in absolute value is down to 56%. Though clearly an improvement on the first panel, a demand elasticity greater than two in absolute value is still highly implausible. The asymmetry between claiming to have very precise prior information about the supply elasticity and having no prior information at all about the demand elasticity remains stark.

What about an analyst who is less confident that the supply elasticity has to be so small? If we instead use an upper bound of $\alpha_0 = 0.075$, 68% posterior confidence bands for the effect of an oil supply shock on price no longer include the Cholesky point estimate, as seen in the

(1,3) panel of Figure 2.

The prior in (33) assumes that we can rule out any value above some threshold α_0 with certainty but regard values right below the threshold as being just as plausible as any other. A more natural representation of prior information is that we are quite confident that the value is small and associate lower probability (but not zero) with larger values. We could do this for example with a Student t distribution with location 0.01, scale parameter 0.03, and truncated to be positive. In the fourth column of Figure 2 we replaced (33) with

$$p(\mathbf{A}) \propto \begin{cases} \left[1+\frac{1}{\nu}\left(\frac{\xi}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} \left[1+\frac{1}{\nu}\left(\frac{\delta}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} \left[1+\frac{1}{\nu}\left(\frac{\beta}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} \left[1+\frac{1}{\nu}\left(\frac{\alpha-c_\alpha}{\sigma_\alpha}\right)^2\right]^{-\frac{\nu+1}{2}} & \text{if } \alpha > 0 \\ 0 & \text{otherwise} \end{cases}$$

with $\nu = 3, \sigma = 100, \sigma_\alpha = 0.03$, and $c_\alpha = 0.01$. Dropping the certainty that the supply elasticity must be below some threshold α_0, even though the density remains quite concentrated around extremely small values, leads to further upward expansion in the likely set of price consequences of an oil supply shock.

3.5 Sources of prior information.

The examples in Figure 2, while generalizing the Cholesky specification, all relied on quite tight prior information that the supply elasticity is very small. Kilian and Murphy (2012) arrived at the bound $\alpha_0 = 0.025$ that we used for illustration in Figure 2 based on analysis of a single historical episode. Caldara, Cavallo, and Iacoviello (2019) generalized Kilian and Murphy's approach to a broader set of historical events and arrived at an instrumental-variables estimate of the short-run supply elasticity of 0.081 with a standard error of 0.037. However, if we relax the confidence in the prior information to be consistent with estimates like these, the posterior credibility sets would widen substantially and we would be left with little basis for drawing structural conclusions.

There is, however, an obvious remedy suggested by Figure 3. In addition to knowledge about the supply elasticity, we also have useful prior information that rules out the extremely large absolute values for the demand elasticity that are implied by all of the examples in Figure 3. By bringing in prior information that the demand elasticity must be negative and is unlikely to be large in absolute value, we can compensate for some of the uncertainty that is introduced when we relax the confidence in the prior information about the supply elasticity. We would argue that the extreme asymmetry represented by the Cholesky example – exact information about the supply elasticity, no information at all about the demand elasticity – is essentially an artifact to which economists have been led by the need to come up with identifying restrictions. In practice, the motivation for a specification like (24) is all too often, "I need three zeroes, and well, here are three."

The Bayesian approach to identification formulates a prior based not on what we "need to assume", but instead on what we know, and importantly, what we don't know. The conclusions from economic models and previous data sets should be represented using priors that are tightly concentrated for magnitudes for which we have very good evidence but with larger variances for magnitudes about which we are less certain. We would argue for bringing in inexact prior information from multiple sources rather than claiming to have exact prior knowledge about a few parameters. Baumeister and Hamilton (2019) illustrated how prior information from multiple sources can be used to inform a four-equation model of the world oil market that includes a role for oil inventories.

We would also like to comment on the role of parameterization. We have represented the structural model in terms of values of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ in a system of equations of the form of (4). Other researchers often parameterize the structural model in terms of the impacts of structural shocks $\mathbf{H}$ as in (18). Uhlig (2017) has argued that the $\mathbf{H}$ parameterization is to be preferred since from the perspective of policy, what we often care about are the equilibrium effects of possible interventions. None of the points we have been making depend on whether the structural model is parameterized as (4) or (18). Whether one is interested in $\mathbf{A}$ or $\mathbf{H}$, the principles are the same and the method of estimation is the same. Prior information about $\mathbf{A}$ and $\mathbf{D}$ can be translated into exactly equivalent prior information about $\mathbf{H}$ using equation (19) and the change-of-variables formula for densities. Bayesian inference, whether about $\mathbf{H}$ or about $\mathbf{A}$, will be the same regardless of parameterization whenever the same prior information used for the different parameterizations reflects the same economic content.

The issue is not whether elements of $\mathbf{A}$ or elements of $\mathbf{H}$ are the objects of interest. The real question is, what are the structural objects about which the researcher has prior information? We would argue that prior information often comes in the form of insights about $\mathbf{A}$ rather than $\mathbf{H}$. Most microfounded models take the form of a system like (4), in which individual equations represent the actions of different agents such as consumers, firms, or government, rather than knowledge about the general equilibrium impacts of the actions of individual actors. Formulating a prior $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$ typically involves looking at previous findings about: elasticities (Baumeister and Hamilton (2019), Brinca et al. (2021), Aastveit et al. (forthcoming), Valenti et al. (2020)); policy rules (Baumeister and Hamilton (2018), Nguyen (2019), Belongia and Ireland (2021)); behavioral equations from economic theory (Aruoba et al. (forthcoming), Lukmanova and Rabitsch (2021)); and responses of agents to permanent changes (Baumeister and Hamilton (2015)). Typically these are most naturally represented as information about $\mathbf{A}$, not $\mathbf{H}$, even though they all have implications for prior information about $\mathbf{H}$.

Notwithstanding, researchers may also have some useful information about the equilibrium impacts of structural shocks. For example, extremely large impacts of policy changes on broad macroeconomic variables may be regarded as unlikely, or we may claim to know a priori the signs of certain elements of **H.** There is no problem in incorporating information about **H** as a supplement to information about **A**. Suppose that for the system given by (1)-(3) (and the necessary implication of those three equations in the form of expression (22)) we had prior information about both the price elasticity of supply $p_1(\alpha)$ and the impact effect of a supply shock on income $p_2(|\mathbf{A}|^{-1}(-\psi - \beta\xi))$. Then we could use the product $p(\mathbf{A}) \propto p_1(\alpha)p_2(|\mathbf{A}|^{-1}(-\psi - \beta\xi))$ as a composite prior for **A**. As discussed by Baumeister and Hamilton (2018), there is no problem with including multiple sources of information about the same parameter, just as there is no problem with using multiple earlier samples that all contain information about a common parameter to form a Bayesian prior in standard settings. The applications in Baumeister and Hamilton (2018, 2019), Grisse (2020), Valenti et al. (2020), Lukmanova and Rabitsch (2021), and Aruoba et al. (forthcoming) all incorporate prior information about both **A** and $\mathbf{A}^{-1}$. Baumeister and Derdzyan (2021) illustrated the use of conditional prior distributions to model the dependence of prior information about different parameters.[12]

A very promising development in structural interpretation of VARs is the use of external instruments developed by Stock and Watson (2012, 2018) and Mertens and Ravn (2013). Here the prior information that the econometrician is relying on is the assumption that the instrument is correlated with a structural shock of interest and uncorrelated with the other structural shocks. For example, Gertler and Karadi (2015) suggested that the change in interest rates within 30 minutes of a policy announcement by the Federal Reserve is correlated only with a shock to monetary policy. Although this is an attractive idea, Campbell et al. (2012) and Nakamura and Steinsson (2018) presented evidence that the Fed announcements are also revealing information about economic fundamentals, which would make them correlated with other structural shocks. Just as with any other identifying assumptions, one could consider a Bayesian generalization of the usual approach to using instrumental variables in structural VARs, as represented by the prior information that the correlation between the instrument and other structural shocks is likely to be small, although the analyst may not be 100% certain that the correlation is zero. Nguyen (2019) provided an illustration of this approach.

4 Set identification using sign restrictions.

The concerns we have raised about traditional approaches to identification such as Cholesky are widely shared by many researchers today. This led Uhlig (2005) and Rubio-Ramírez,

[12] Kilian and Zhou (2020) asserted that "Baumeister and Hamilton's approach is not designed to handle the restrictions on $[\mathbf{A}^{-1}]$ typical of conventional oil market models, except in the special case of a recursively identified model." The same claim was repeated in Kilian and Lütkepohl (2017, p. 454). The statement is false. The method described by Baumeister and Hamilton (2018, 2019) for incorporating information about both **A** and $\mathbf{A}^{-1}$ is completely general.

Waggoner, and Zha (2010) to develop approaches that rely not on zero restrictions on $\mathbf{A}$ or $\mathbf{H}$ but instead on prior knowledge about the signs of effects of structural shocks. In the oil-market example, Kilian and Murphy (2012) assumed that the signs of impacts are given by

$$\text{sign}(\mathbf{H}) = \begin{bmatrix} - & + & + \\ - & + & - \\ + & + & + \end{bmatrix}. \tag{34}$$

The first column of (34) says that an oil supply shock lowers production, lowers economic activity, and raises the price of oil. The second column indicates the effects of shocks to economic activity and the third column the effects of shocks to oil demand.

4.1 Description of the Rubio-Ramírez, Waggoner, and Zha (2010) algorithm.

The Rubio-Ramírez, Waggoner, and Zha (2010) algorithm for estimating structural impulse-response functions $\mathbf{\Psi}_s\mathbf{H}$ using sign restrictions for identification proceeds as follows. We generate a draw for $\mathbf{\Omega}$ from an inverse-Wishart distribution with scale matrix $T\hat{\mathbf{\Omega}}_{MLE}$ and $T-mn-1$ degrees of freedom, and use this $\mathbf{\Omega}$ to generate a draw for vec($\mathbf{\Pi}'$) from a $N(\text{vec}(\hat{\mathbf{\Pi}}'),$ $\mathbf{\Omega} \otimes [\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}_{t-1}'])$ distribution. Note that this is exactly the same distribution that was used in Section 2.4 to calculate standard errors for the IRF. There we noted that the distribution can be motivated either as an approximation to the frequentist distribution of the reduced-form parameters $(\mathbf{\Omega}, \mathbf{\Pi})$ or as a characterization of the Bayesian posterior distribution of these parameters if the researcher began with an uninformative prior in the sense of Jeffreys.[13] Next the researcher draws an $(n \times n)$ matrix $\mathbf{Q}$ that is calculated from a QR decomposition of an $(n \times n)$ matrix of independent $N(0, 1)$ variables that the researcher generates. The QR decomposition means that every generated value for $\mathbf{Q}$ is an orthonormal matrix ($\mathbf{Q}'\mathbf{Q} = \mathbf{I}_n$). Baumeister and Hamilton (2015) showed that the (i, j) element of $\mathbf{Q}$ generated by this algorithm has a density given by

$$p(q_{ij}) = \begin{cases} \frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)}(1-q_{ij}^2)^{(n-3)/2} & \text{if } q_{ij} \in [-1, 1] \\ 0 & \text{otherwise} \end{cases}. \tag{35}$$

Let $\mathbf{P}$ denote the Cholesky factor of a generated draw for $\mathbf{\Omega}$. Researchers then calculate $\mathbf{H} = \mathbf{P}\mathbf{Q}'$ and check whether this proposed value for $\mathbf{H}$ satisfies the desired sign restrictions such as those in (34). If it does, the draw is retained, and if it does not, then the draw is rejected and a new draw is generated. Researchers then typically report the median of the retained set of values for the row i column j element of $\mathbf{\Psi}_s\mathbf{H}$, which we denote ζ_{ijs}, as if it

[13] One could also use more informative priors for the reduced-form coefficients such as the Minnesota prior, but the vast majority of applications follow Uhlig (2005) in using the Jeffreys prior.

was a point estimate of the effect of a one-standard deviation shock to structural disturbance j on the ith element of $\mathbf{y}$ after s periods, and report 68% or 95% of the set of retained draws for ζ_{ijs} as if they represented a confidence interval.

4.2 A frequentist critique of the typical sign-restricted VAR application.

If the only prior information available to a frequentist was information about signs as in (34), the structural model would only be set-identified. This means that there is a set of possible values for ζ_{ijs} that are all associated with the single value for $(\hat{\mathbf{\Omega}}, \hat{\mathbf{\Pi}})$ that maximizes the likelihood function and that are all consistent with all the prior information that the econometrician has about the structure. One simple algorithm[14] for finding the upper and lower bounds for this set is to use the fixed MLE $\hat{\mathbf{\Pi}}$ to calculate $\hat{\mathbf{\Psi}}_s$, the fixed MLE $\hat{\mathbf{\Omega}}$ to find its Cholesky factor $\hat{\mathbf{P}}\hat{\mathbf{P}}' = \hat{\mathbf{\Omega}}$, and use the algorithm in Section 4.1 to generate a very large number of draws of $\mathbf{Q}$. We calculate ζ_{ijs} from the (i, j) element of $\hat{\mathbf{\Psi}}_s\hat{\mathbf{P}}\mathbf{Q}'$ for each draw of $\mathbf{Q}$ and calculate the maximum and minimum of values of ζ_{ijs} across *all* of the generated draws (not just 95% of the generated draws). As the number of draws of $\mathbf{Q}$ goes to infinity these would would converge to the range of values of ζ_{ijs} that are all consistent with the maximum-likelihood estimates of the reduced-form parameters. Denote these maximum-likelihood estimates of the boundaries of the set of possible values for ζ_{ijs} by $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$. A frequentist would want further to account for uncertainty about the estimation of $\mathbf{\Pi}$ and $\mathbf{\Omega}$, resulting in a confidence set that is strictly broader than $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$ as described for example by Gafarov, Meier, and Montiel Olea (2018). Moon and Schorfheide (2012) and Watson (2019) sharply criticized the practice of reporting 68% or 95% error bands that are strictly smaller than $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$.

4.3 A Bayesian critique of the typical sign-restricted VAR application.

Every value for the structural impulse response ζ_{ijs} that is in the set $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$ maximizes the likelihood function of the observed data and is consistent with all the restrictions implied by (34). If a Bayesian were to report that the values $\hat{\underline{\zeta}}_{ijs}$ or $\hat{\bar{\zeta}}_{ijs}$ are relatively unlikely and fall outside of a 68% or 95% credibility set, the basis for this exclusion must be more than something that is observed in the data and more than the information embodied in the sign restrictions (34). One possible justification for doing this would be if the distribution in (35) was interpreted as Bayesian prior information, specifically, information the econometrician

[14] For a discussion of this and alternative algorithms to estimate $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$ see Giacomini, Kitagawa, and Read (2021).

had about the structural model before seeing the data and in addition to the information represented by (34).

The usual motivation for the distribution in (35) is that it can be viewed as a uniform measure over the set of orthonormal matrices. The argument is that because this measure weights all orthonormal matrices equally, it does not influence the statistics that the Bayesian reports. This does not answer the question of the source of the information that led us to conclude that $\underline{\hat{\zeta}}_{ijs}$ or $\hat{\bar{\zeta}}_{ijs}$ are relatively unlikely values for ζ_{ijs}. If that conclusion did not come from prior information about $\mathbf{Q}$, then from where did it come? The distribution assumed for $\mathbf{Q}$ incontrovertibly played a role in the conclusions reported if it led a researcher to report that $\underline{\hat{\zeta}}_{ijs}$ is outside the set of values that are plausible given the data.

A separate issue is what one means by the claim that a distribution puts equal weight on all the possibilities. When $n = 2$, the set of orthornormal matrices can be indexed by an angle of rotation or reflection θ.[15] The distribution in (35) implies that all angles are equally likely, that is, a distribution for θ that is uniform over $(0,2\pi)$. If the researcher intends to draw an inference about this angle θ, then the prior could correctly be said to weight all possibilities equally.

But a distribution that is uninformative about one function of the parameters (such as the angle of rotation θ) is of necessity informative about other functions of parameters (such as the value of $q_{ij} = \cos(\theta)$). Nobody reports the angle of rotation associated with $\mathbf{Q}$ as if it were a structural magnitude of interest. Instead applied researchers report for example elements of $\mathbf{PQ}'$, which are interpreted as the impacts of one-standard-deviation structural shocks. From the (1,1) element of $\mathbf{PQ}'$ we see that the effect on variable 1 of a one-standard deviation shock to the first structural equation is given by $p_{11}q_{11}$.

Figure 4 shows what the distribution of individual elements of $\mathbf{Q}$ looks like. If there are $n = 3$ variables in the VAR, then all values in $(-1, 1)$ are equally likely. By contrast, when $n = 2$, values near ± 1 are more likely, and when $n > 3$ values near zero are more likely.

When $n = 2$, the procedure amounts to prior knowledge that the effect of the first structural shock on the first variable is more likely to be large than small. What is the basis for this prior information? How is it something we know before seeing the data, regardless of what kind of data we consider, and regardless of the economic content of the first structural equation? Users of the sign-restriction methodology never defend such an interpretation, never explain the basis for excluding values like $\underline{\hat{\zeta}}_{ijs}$ as implausible, and typically do not claim to have relied on any prior information other than the information about the signs of impacts. In the absence of prior information about $\mathbf{Q}$, there is no basis for reporting a point estimate of a structural magnitude that is only set-identified, and no basis for claiming that we have 95% confidence

[15] Any orthonormal (2×2) matrix $\mathbf{Q}$ can be written as either $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ or $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ for some $\theta \in (0, 2\pi)$.

that the true value lies in some subset of the draws generated by the algorithm.

Giacomini and Kitagawa (2021) advocated what they described as "robust Bayesian inference." They proposed considering the range of posterior means that could be arrived at from *some* prior distribution for $\mathbf{Q}$, and showed that asymptotically this is the same set as $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$. In accounting for uncertainty in estimation of $(\mathbf{\Omega}, \mathbf{\Pi})$ from a robust Bayesian perspective, they developed an algorithm that again results in reporting a set that is strictly broader than $(\hat{\underline{\zeta}}_{ijs}, \hat{\bar{\zeta}}_{ijs})$.

To summarize, if the claim of the researcher is to have used no prior information other than the signs of certain magnitudes, we see no justification for either a Bayesian or a frequentist to report the median, mode, or any other summary statistic of the retained draws as if it were an optimal estimate of some magnitude of interest, nor to report 68% or 95% of the retained draws around the median as if that range summarizes our confidence in the estimate. We recommend that if a researcher is implicitly relying on prior information (as anyone who reports posterior medians or modes or 68% or 95% credibility sets implicitly is), it is desirable to state and defend this prior information.

4.4 Computational considerations.

Here we highlight another aspect of sign identification as typically implemented: as researchers impose more restrictions, the number of accepted draws shrinks. For example, Kilian and Murphy (2014) imposed sign restrictions as well as the restriction that the supply elasticity had to be lie in $(0, 0.025)$ and employed a number of other criteria for discarding values of $\mathbf{H}$ generated by the algorithm. The code for their paper posted at the *Journal of Applied Econometrics* data archive generates 5 million draws for the vector of possible parameters of which only 16 satisfy all the authors' criteria.

Uhlig (2017) argued that when so many draws are rejected, the identification is sharp and that this is a good thing. We have several concerns about this. The first is the question we discussed in Section 3.4, which is whether restrictions such as the claim that the supply elasticity must be less than 0.025 are completely credible. Second is the practical issue of what to conclude from the 16 retained draws, and whether for example these form an adequate basis for making statements about the identified set or other magnitudes of interest.

Kilian and Murphy (2014) did not offer a Bayesian interpretation of their procedure. The method that they used in their paper to report a single estimate from the set of 16 retained draws was to select the draw that had an "impact price elasticity of oil demand in use closest to the posterior median of that elasticity among the admissible structural models" (p. 464). Results from running their code as publicly posted, which incorporates this criterion for selecting a representative draw, are plotted as the dotted red lines in our Figure 5. These show the effects of what Kilian and Murphy (2014) called a speculative demand shock on their measure of real economic activity and on the real price of oil. This figure reproduces two of the panels shown in Figure 1 of their article. A researcher who ran this code and looked at this output might describe the findings as Kilian and Murphy did on pages 464-465:

> a positive speculative demand shock is associated with an immediate jump in the real price of oil. The real price response overshoots, before declining gradually. The effects on global real activity and global oil production are largely negative, but small.

We reran their posted code making only one change. In the original code, the seed used in the random number generator is 316. We reran the same code using instead a seed for the random number generator of 613. The blue solid lines in Figure 5 show the structural estimates that result when this different seed for the random number generator is used. A researcher who ran their code using a random number seed of 613 instead of 316 might describe the findings as follows:

> a positive speculative demand shock is associated with an immediate large drop in economic activity and a small positive effect on price.

Using a random number seed of 613 thus leads to completely different policy implications compared to a seed of 316.

By contrast, the results in Baumeister and Hamilton (2018, 2019) are all based on 1 million retained draws. This allows us to characterize accurately the posterior distribution that results from explicit prior structural information and use a standard loss function to summarize any properties of interest of this known distribution.

5 Conclusion.

Whether the goal is applied research or policy guidance, there is a clear answer to the question of how to use a vector autoregression to estimate any structural magnitude of interest. The likelihood function summarizes everything the data could tell us about parameters. If our prior information about the structure is exact, we can use maximum-likelihood estimation to obtain estimates that are asymptotically optimal from a frequentist perspective and are invariant with respect to how the model is parameterized. If the prior information is inexact, we should summarize that prior information in the form of a probability distribution and use Bayes' Law to characterize uncertainty about structural magnitudes that remains after observing the data.

References

Aastveit, Knut Are, Hilde C. Bjørnland, and Jamie L. Cross (forthcoming). "Inflation Expectations and the Pass-Through of Oil Prices," *Review of Economics and Statistics.*

Antolín-Díaz, Juan, and Juan F. Rubio-Ramírez (2018). "Narrative Sign Restrictions for SVARs," *American Economic Review* 108: 2802-2829.

Aruoba, S. Borağan, Marko Mlikota, Frank Schorfheide, and Sergio Villalvazo (forthcoming). "SVARs with Occasionally-Binding Constraints," *Journal of Econometrics.*

Basher, Syed Abul, Alfred A. Haug, and Perry Sadorsky (2018). "The Impact of Oil-Market Shocks on Stock Returns in Major Oil-Exporting Countries," *Journal of International Money and Finance* 86: 264–280.

Baumeister, Christiane, and Aram Derdzyan (2021). "Structural Changes in Energy Markets: A New Approach to Modeling Time-Varying Dynamics," working paper, University of Notre Dame.

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

Baumeister, Christiane, and James D. Hamilton (2018). "Inference in Structural Vector Autoregressions when the Identifying Assumptions Are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations," *Journal of Monetary Economics* 100: 48-65.

Baumeister, Christiane, and James D. Hamilton (2019). "Structural Interpretation of Vector Autoregressions with Incomplete Information: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review* 109: 1873-1910.

Belongia, Michael T., and Peter N. Ireland (2021). "A Classical View of the Business Cycle," *Journal of Money, Credit and Banking* 53: 333-366.

Bernanke, Ben S. (1986). "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy* 25: 49-99.

Brinca, Pedro, Joao B. Duarte, and Miguel Faria e Castro (2021). "Measuring Sectoral Supply and Demand Shocks during COVID-19," *European Economic Review* 139: 103901.

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello (2019). "Oil Price Elasticities and Oil Price Fluctuations," *Journal of Monetary Economics* 103: 1-20.

Campbell, Jeffrey R., Charles L. Evans, Jonas D.M. Fisher, and Alejandro Justiniano (2012). "Macroeconomic Effects of Federal Reserve Forward Guidance," *Brookings Papers on Economic Activity* 43(1): 1-80.

Chao, John C. and Peter C.B. Phillips (2002). "Jeffreys Prior Analysis of the Simultaneous Equations Model in the Case with $n+1$ Endogenous Variables," *Journal of Econometrics* 111: 251-283.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (1999). "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics*, vol.

1, pp. 65-148, edited by John B. Taylor and Michael Woodford, Elsevier.

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

Drèze, Jacques H. (1975). "Bayesian Theory of Identification in Simultaneous Equations Models," in *Studies in Bayesian Econometrics and Statistics*, edited by Stephen E. Fienberg and Arnold Zellner, Amsterdam: North-Holland Publishing Co.

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

Gafarov, Bulat, Matthias Meier, and José Luis Montiel Olea (2018). "Delta-Method Inference for a Class of Set-Identified SVARs," *Journal of Econometrics* 203: 316-327.

Gertler, Mark, and Peter Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics* 7: 44-76.

Giacomini, Raffaella, and Toru Kitagawa (2021). "Robust Bayesian Inference for Set-Identified Models," *Econometrica* 89: 1519-1556.

Giacomini, Raffaella, Toru Kitagawa, and Matthew Read (2021). "Robust Bayesian Analysis for Econometrics," working paper, University College London.

Grisse, Christian (2020). "The Effect of Monetary Policy on the Swiss Franc: An SVAR Approach," Swiss National Bank working paper 2020-02.

Güntner, Jochen H.F. (2014), "How Do Oil Producers Respond to Demand Shocks?" *Energy Economics* 44: 1-13.

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

Herrera, Ana Maria, and Sandeep Rangaraju (2020). "The Effect of Oil Supply Shocks on US Economic Activity: What Have We Learned?", *Journal of Applied Econometrics* 35: 141-159.

Inoue, Atsushi, and Lutz Kilian (forthcoming). "Joint Bayesian Inference about Impulse Responses in VAR Models," *Journal of Econometrics.*

Jeffreys, Harold (1946). "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences*, 186: 453-461.

Karlsson, Sune (2013). "Forecasting with Bayesian Vector Autoregression," *Handbook of Economic Forecasting,* Volume 2B, pp. 791-897, edited by Graham Elliott and Allan Timmermann, Elsevier.

Kilian, Lutz (2009). "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review* 99: 1053-1069.

Kilian, Lutz, and Helmut Lütkepohl (2017). *Structural Vector Autoregressive Analysis*, Cambridge 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.

Kilian, Lutz, and Daniel P. Murphy (2014). "The Role of Inventories and Speculative Trading in the Global Market for Crude Oil," *Journal of Applied Econometrics* 29: 454-478.

Kilian, Lutz, and Xiaoqing Zhou (2020). "The Econometrics of Oil Market VAR Models," CEPR working paper 14460.

Kleibergen, Frank, and Eric Zivot (2003). "Bayesian and Classical Approaches to Instrumental Variable Regression," *Journal of Econometrics* 114: 29-72.

Lindley, Dennis Victor (1972). *Bayesian Statistics: A Review*, Society for Industrial and Applied Mathematics.

Ludvigson, Sydney C., Sai Ma, and Serena Ng (2017). "Shock Restricted Structural Vector-Autoregressions," NBER Working Paper 23225.

Lukmanova, Elizaveta and Katrin Rabitsch (2021). "New Evidence on Monetary Transmission: Interest Rate Versus Inflation Target Shocks," working paper, KU Leuven.

Mertens, Karel and Morten O. Ravn (2013). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review* 103: 1212-1247.

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

Nakamura, Emi, and Jón Steinsson (2018). "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics* 133: 1283-1330.

Nguyen, Lam (2019). "Bayesian Inference in Structural Vector Autoregression with Sign Restrictions and External Instruments," working paper, UCSD.

Phillips, Peter C.B. (1991). "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics* 6: 334-364.

Plagborg-Møller, Mikkel, and Christian K. Wolf (2021). "Local Projections and VARs Estimate the Same Impule Responses," *Econometrica* 89: 955-980.

Riggi, Marianna, and Fabrizio Venditti (2015). "The Time Varying Effect of Oil Price Shocks on Euro-area Exports," *Journal of Economic Dynamics and Control* 59: 75-94.

Rothenberg, Thomas J. (1973). *Efficient Estimation with A Priori Information*, Yale University Press.

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

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.

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

Stock, James H., and Mark W. Watson (2012). "Disentangling the Channels of the 2007-09 Recession," *Brookings Papers on Economic Activity* 43(1): 81–156.

Stock, James H., and Mark W. Watson (2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal* 128: 917-948.

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.

Uhlig, Harald (2017). "Shocks, Sign Restrictions, and Identification," in *Advances in Economics and Econometrics* Volume 2, pp. 95-127, edited by Bo Honoré, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, Cambridge University Press.

Valenti, Daniele, Matteo Manera, and Alessandro Sbuelz (2020). "Interpreting the Oil Risk Premium: Do Oil Price Shocks Matter?," *Energy Economics* 91: 104906.

Watson, Mark (2019). "Comment on 'On the Empirical (Ir)Relevance of the Zero Lower Bound' by D. Debortoli, J. Gali, and L. Gambetti," *NBER Macroeconomics Annual.*

Zhou, Xiaoqing (2020). "Refining the Workhorse Oil Market Model," *Journal of Applied Econometrics* 35: 130-140.

Appendix A: Derivation of equation (11).

Expression (10) can be rewritten

$$\left[\begin{array}{cc} \sum_{t=1}^{T}\hat{u}_t^s\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{u}_t^s\hat{\varepsilon}_t^p \\ \sum_{t=1}^{T}\hat{u}_t^y\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{u}_t^y\hat{\varepsilon}_t^p \end{array}\right]\left[\begin{array}{c} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{array}\right] = \left[\begin{array}{c} \sum_{t=1}^{T}\hat{u}_t^s\hat{\varepsilon}_t^q \\ \sum_{t=1}^{T}\hat{u}_t^y\hat{\varepsilon}_t^q \end{array}\right]$$

$$\begin{aligned} &\left[\begin{array}{ccc} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{array}\right]\left[\begin{array}{cc} \sum_{t=1}^{T}\hat{\varepsilon}_t^q\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{\varepsilon}_t^q\hat{\varepsilon}_t^p \\ \sum_{t=1}^{T}\hat{\varepsilon}_t^y\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{\varepsilon}_t^y\hat{\varepsilon}_t^p \\ \sum_{t=1}^{T}\hat{\varepsilon}_t^p\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{\varepsilon}_t^p\hat{\varepsilon}_t^p \end{array}\right]\left[\begin{array}{c} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{array}\right] \\ &= \left[\begin{array}{ccc} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{array}\right]\left[\begin{array}{c} \sum_{t=1}^{T}\hat{\varepsilon}_t^q\hat{\varepsilon}_t^q \\ \sum_{t=1}^{T}\hat{\varepsilon}_t^y\hat{\varepsilon}_t^q \\ \sum_{t=1}^{T}\hat{\varepsilon}_t^p\hat{\varepsilon}_t^q \end{array}\right] \end{aligned}$$

$$\left[\begin{array}{ccc} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{array}\right]\left[\begin{array}{ccc} \sum_{t=1}^{T}\hat{\varepsilon}_t^q\hat{\varepsilon}_t^q & \sum_{t=1}^{T}\hat{\varepsilon}_t^q\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{\varepsilon}_t^q\hat{\varepsilon}_t^p \\ \sum_{t=1}^{T}\hat{\varepsilon}_t^y\hat{\varepsilon}_t^q & \sum_{t=1}^{T}\hat{\varepsilon}_t^y\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{\varepsilon}_t^y\hat{\varepsilon}_t^p \\ \sum_{t=1}^{T}\hat{\varepsilon}_t^p\hat{\varepsilon}_t^q & \sum_{t=1}^{T}\hat{\varepsilon}_t^p\hat{\varepsilon}_t^y & \sum_{t=1}^{T}\hat{\varepsilon}_t^p\hat{\varepsilon}_t^p \end{array}\right]\left[\begin{array}{c} 1 \\ -\hat{\delta}_{IV} \\ -\hat{\beta}_{IV} \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \qquad (36)$$

from which (11) follows from the definitions of $\mathbf{\Gamma}, \boldsymbol{\eta}$, and $\hat{\mathbf{\Omega}}$.

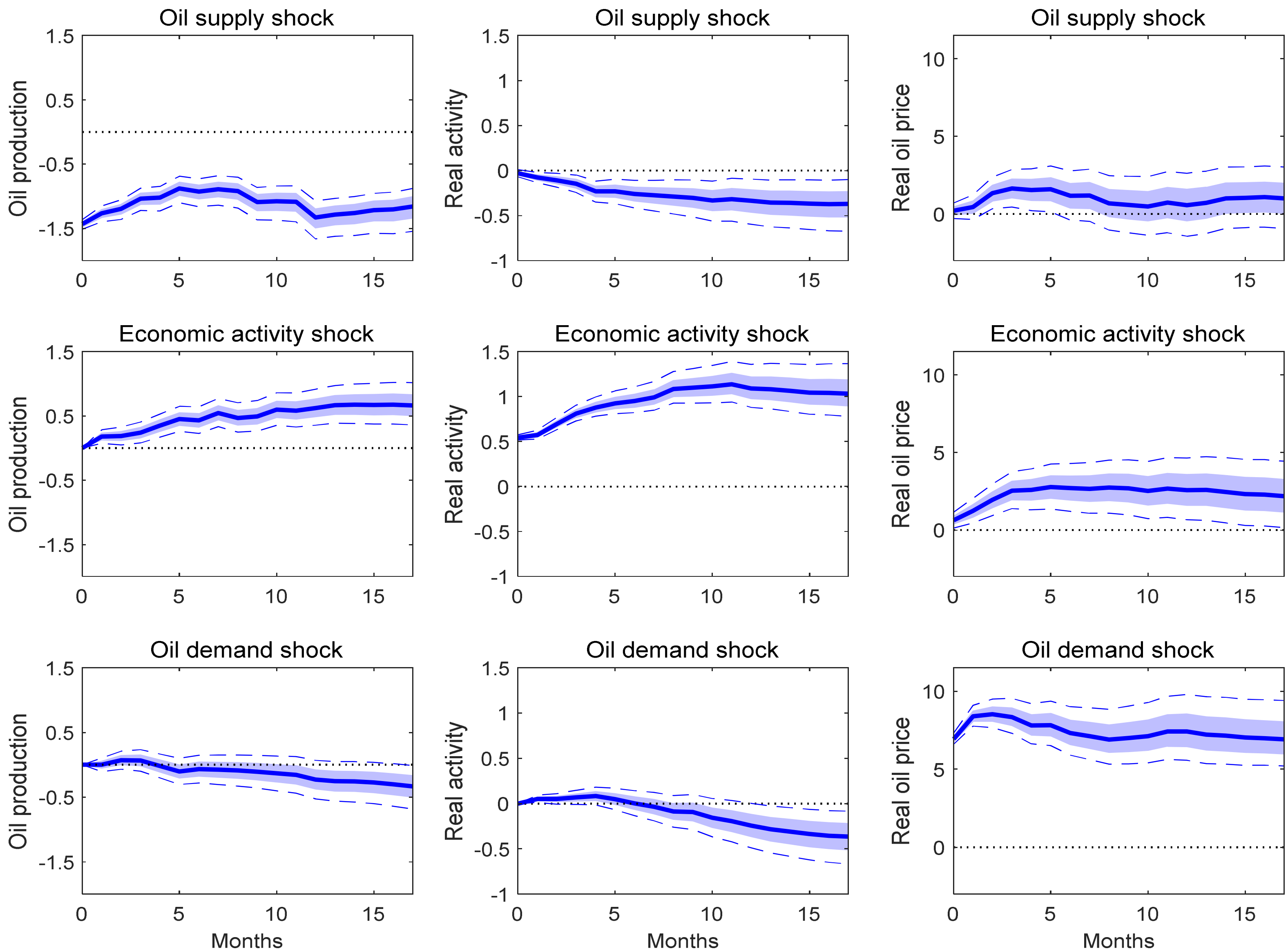

Figure 1. Impulse-response functions for 3-variable oil market model under traditional Cholesky identification. Solid blue lines: maximum-likelihood estimate; shaded regions: 68% error bands; dashed blue lines: 95% error bands.

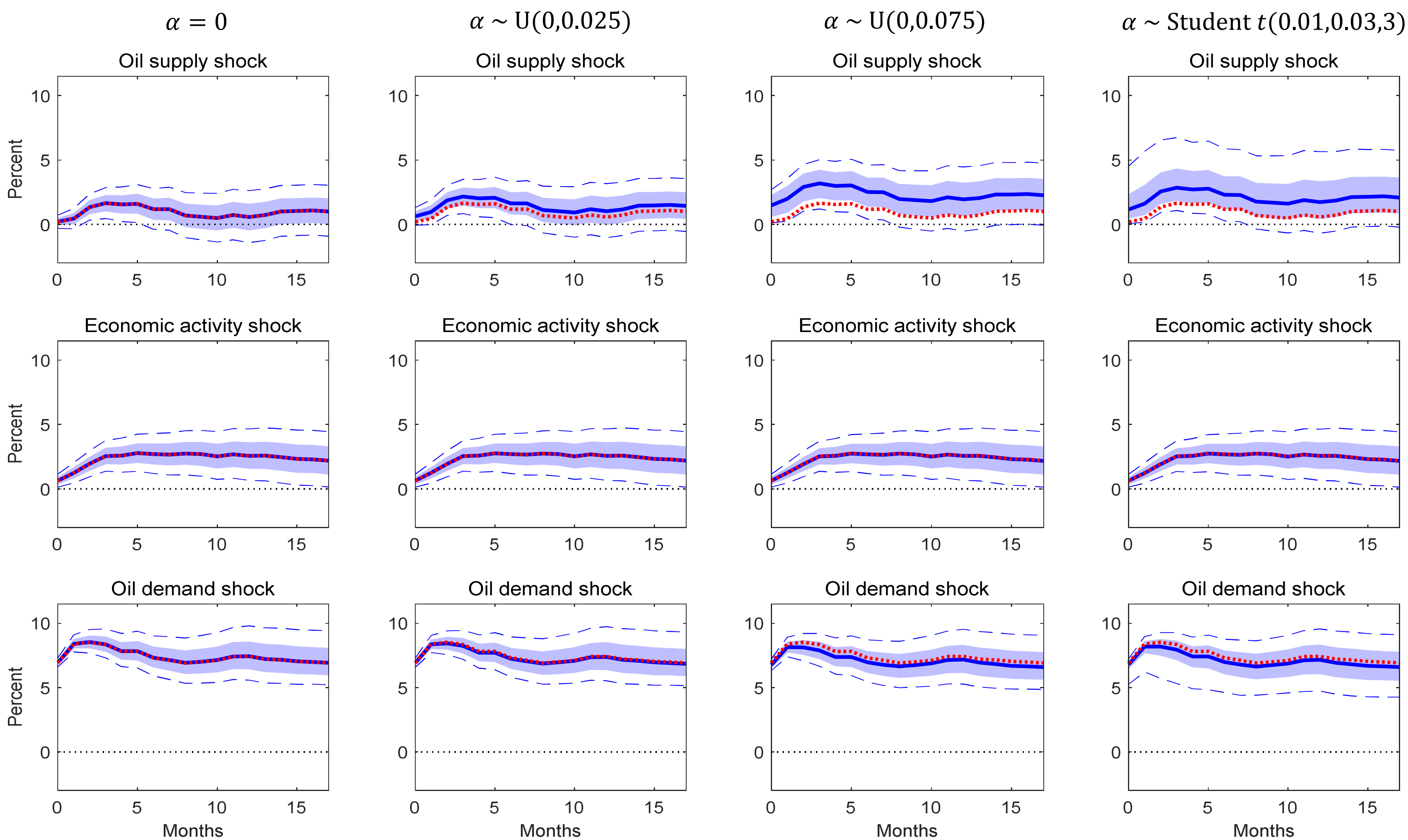

Figure 2. Responses of the real price of oil after an oil supply shock, an aggregate demand shock and an oil-specific demand shock under different priors for the short-run price elasticity of supply as indicated in the header. Red dotted lines: maximum-likelihood estimates under Cholesky identification; blue solid lines: median of Bayesian posterior distribution; shaded regions: 68% posterior credibility set; blue dotted lines: 95% posterior credibility set.

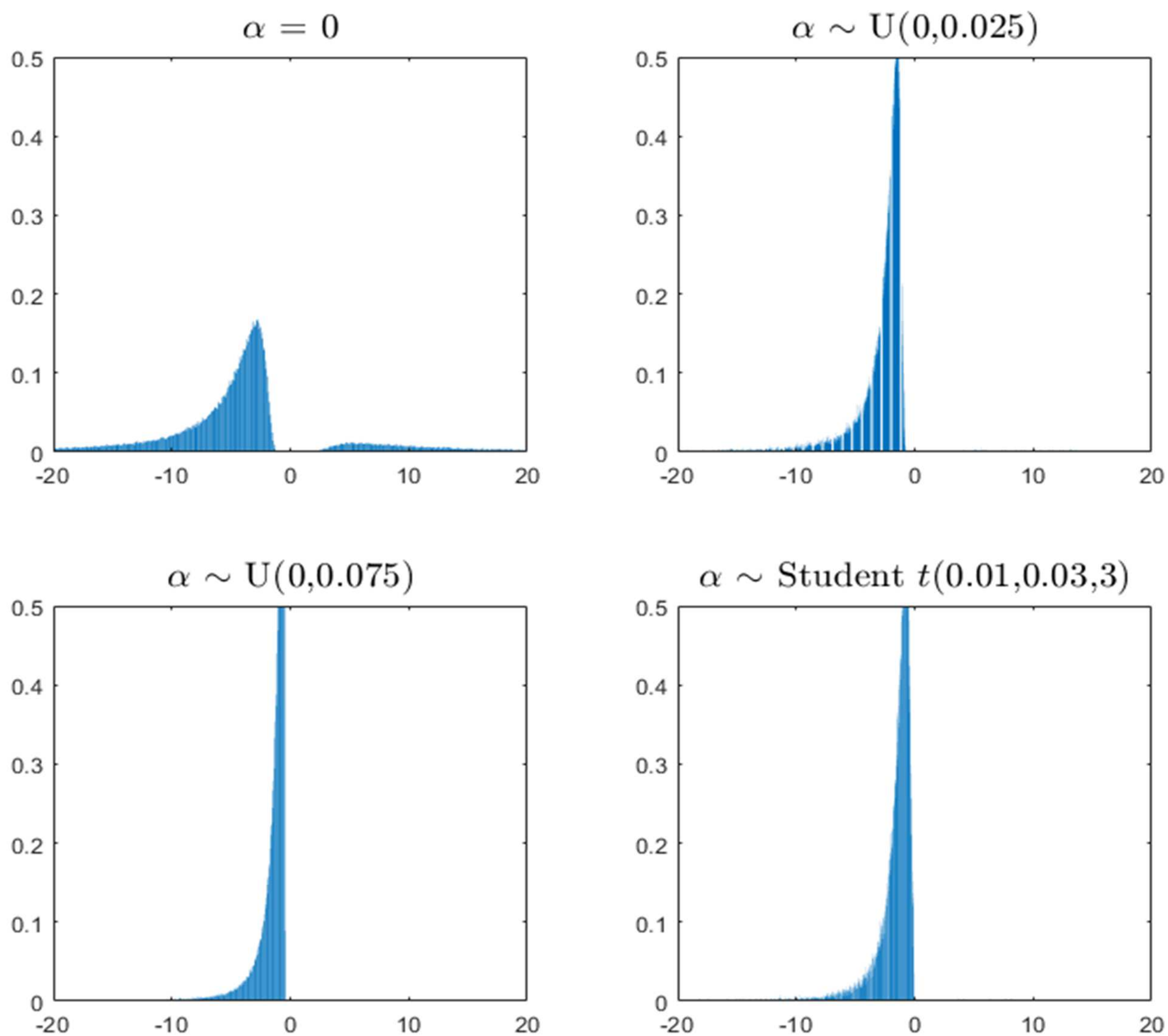

Figure 3. Posterior distributions of the short-run price elasticity of oil demand under different priors for the oil supply elasticity as indicated in the headers.

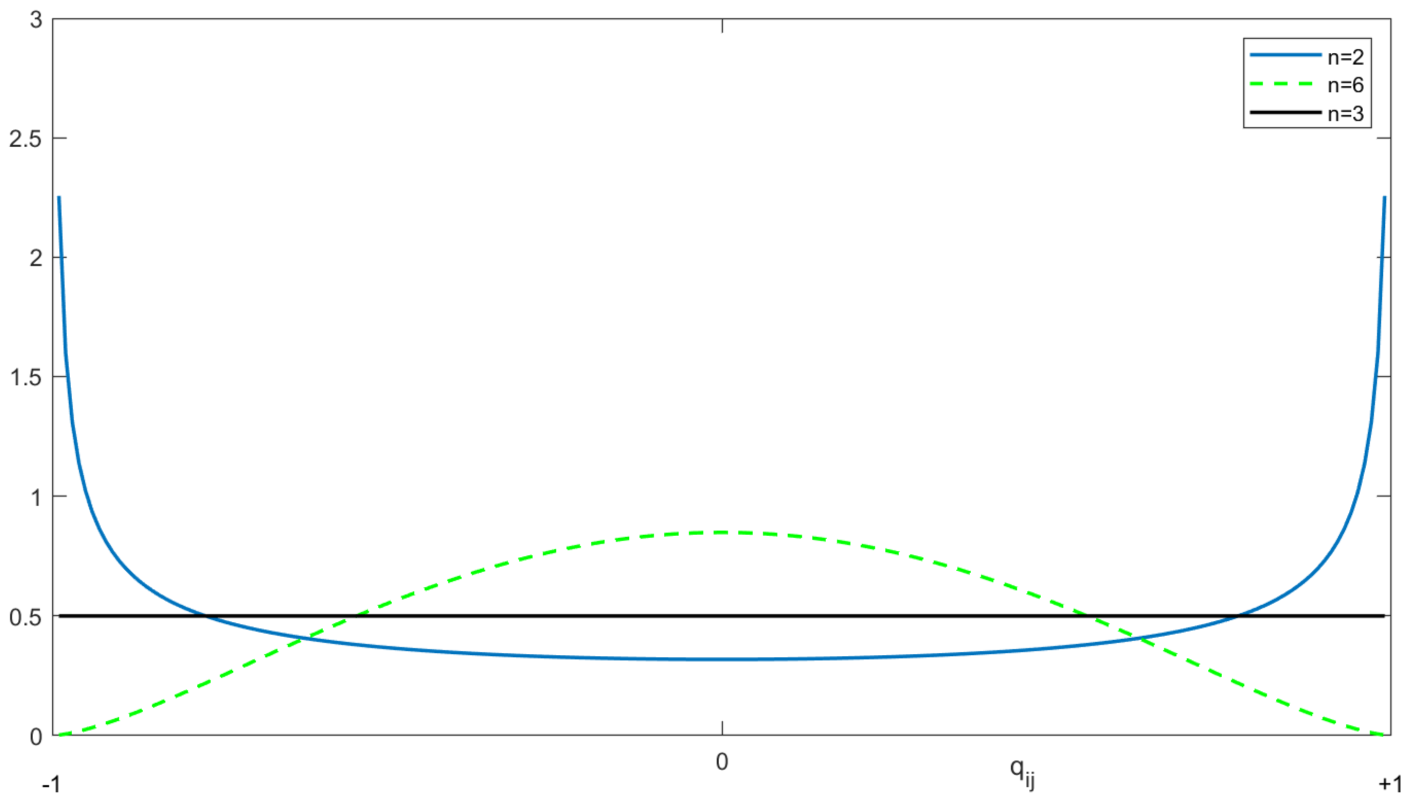

Figure 4. Distribution of elements of the matrix **Q** generated by the sign-restriction algorithm for three different values of *n*.

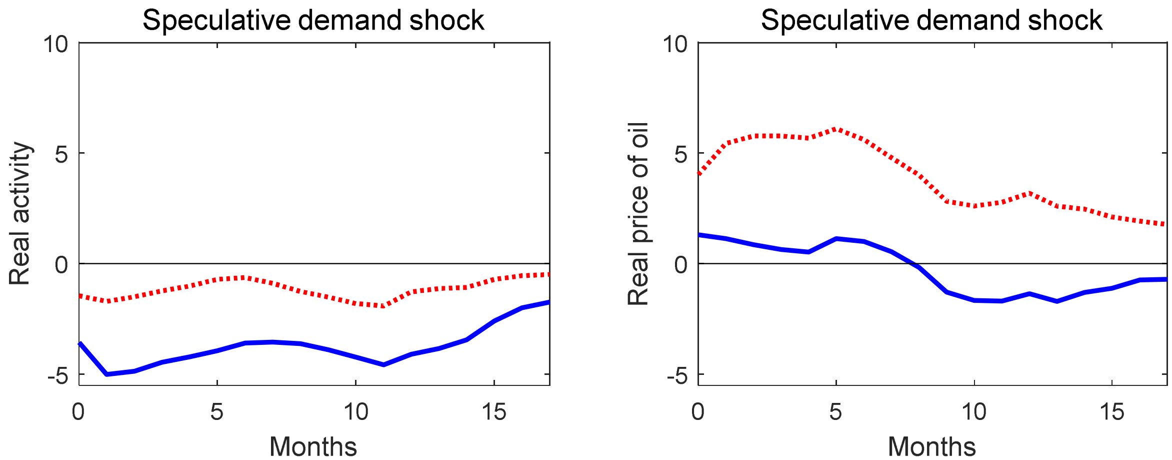

Figure 5. Effects of speculative oil demand shock for the Kilian and Murphy (2014) specification and data set using two different seeds for the random number generator. Left panel: effect on real activity. Right panel: effect on real price of oil. Red dotted lines: seed = 316, which was the original seed used by Kilian and Murphy (2014) and which reproduces panels (3,2) and (3,3) in Kilian and Murphy's Figure 1. Blue solid lines: seed = 613.