NBER WORKING PAPER SERIES

ADVANCES IN STRUCTURAL VECTOR AUTOREGRESSIONS WITH IMPERFECT IDENTIFYING INFORMATION

Christiane Baumeister
James D. Hamilton

Working Paper 27014
http://www.nber.org/papers/w27014

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

We thank Diego Kaenzig and Lam Nguyen for helpful comments on an earlier draft of this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Advances in Structural Vector Autoregressions with Imperfect Identifying Information
Christiane Baumeister and James D. Hamilton
NBER Working Paper No. 27014
April 2020
JEL No. C11,C32,Q43

ABSTRACT

This paper examines methods for structural interpretation of vector autoregressions when the identifying information is regarded as imperfect or incomplete. We suggest that a Bayesian approach offers a unifying theme for guiding inference in such settings. Among other advantages, the unified approach solves a problem with calculating elasticities that appears not to have been recognized by earlier researchers. We also call attention to some computational concerns of which researchers who approach this problem using other methods should be aware.

Christiane Baumeister
Department of Economics
University of Notre Dame
3028 Jenkins Nanovic Hall
Notre Dame, IN 46556
and NBER
cbaumeis@nd.edu

James D. Hamilton
Department of Economics, 0508
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER
jhamilton@ucsd.edu

1 Introduction.

The problem of identification – drawing causal or structural conclusions from the correlations we observe in the data – is often the core challenge of empirical economic research. The traditional approach to identification is to bring in additional information in the form of identifying assumptions, such as restrictions that certain magnitudes have to be zero. Although this approach is very common in empirical economic studies, it would be hard to find an economic researcher who does not entertain some doubts about whether the identifying restrictions are really valid.

These doubts have led to substantial interest among empirical macroeconomists in methods that make use of incomplete identifying assumptions, such as restrictions on the signs rather than the magnitudes of certain parameters of interest. These methods have been applied to interpreting vector autoregressions in hundreds of published studies.[1] Surveys of the literature have been provided by Kilian and Lütkepohl (2017), Uhlig (2017), and Kilian and Zhou (2020).

In this paper we propose a unifying principle for approaching these questions that much of the literature has overlooked. We suggest that what are usually thought of as identifying assumptions should more generally be described as information that the analyst had about the economic structure before seeing the data. We maintain that such information is most naturally represented as a Bayesian prior distribution over certain features of the economic structure. Traditional point identification can be viewed as a special case of a dogmatic Bayesian prior – values that we knew for certain before we saw the data. The natural way to acknowledge that our prior information about the structure is less than perfect is to reduce the confidence reflected in those prior distributions. Application of Bayes' Law will then result in those doubts about the true structure being incorporated in the conclusions we draw after having seen the data.

In Section 2 we illustrate this theme using a number of earlier studies and correct some misunderstandings that have appeared in the recent literature. Section 3 demonstrates that a common approach to estimating elasticities in sign-restricted vector autoregressions, used for example by Kilian and Murphy (2012, 2014), Güntner (2014), Riggi and Venditti (2015), Kilian and Lütkepohl (2017), Ludvigson, Ma, and Ng (2017), Antolín-Díaz and Rubio-Ramírez (2018), Basher et al. (2018), Herrera and Rangaraju (2020), and Zhou (2020), is incorrect. We show how a unified Bayesian approach fixes this problem. Section 4 comments on some computational issues of which users of sign-restricted vector autoregressions should be aware. Here we again note an advantage of a unified Bayesian approach in avoiding these problems. Section 5 discusses auxiliary identifying information in the form of instruments or proxy variables. Section 6 briefly concludes.

Our core recommendation is that instead of regarding identifying assumptions as something

[1] See Baumeister and Hamilton (2020a) for a partial list of these.

one simply has to make, researchers should reflect carefully on what we know on the basis of other data sets or economic theory and how confident we are in this knowledge. Both the prior information and our doubts about the quality of the information should be incorporated in any conclusions we draw from the data.

2 Alternative approaches to structural identification.

A wide class of dynamic structural models takes the form

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

Here $\mathbf{y}_t$ is an $(n \times 1)$ vector of observed variables and $\mathbf{x}_{t-1}$ is a $(k \times 1)$ vector containing m lags of $\mathbf{y}$ and a constant term, $\mathbf{x}'_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, ..., \mathbf{y}'_{t-m}, 1)'$, so $k = mn + 1$. This is a system of n structural equations in which $\mathbf{A}$ is an $(n \times n)$ matrix of contemporaneous coefficients and $\mathbf{B}$ captures structural dynamics. The structural shocks $\mathbf{u}_t$ are assumed to be mutually uncorrelated white noise with $E(\mathbf{u}_t\mathbf{u}'_t)$ given by the diagonal matrix $\mathbf{D}$.[2]

Associated with this system is a reduced-form vector autoregression:

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

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

The relation between reduced-form residuals and the structural shocks is

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

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

The reduced-form parameters $\mathbf{\Phi}$ and $\mathbf{\Omega}$ are readily estimated by OLS. However, knowledge of the reduced-form parameters is not enough information to uncover the structural parameters $\mathbf{A}, \mathbf{B}$, and $\mathbf{D}$ of interest. Drawing an inference about the structural parameters requires bringing in additional information about the structure from other sources.

2.1 The traditional approach to identification.

The traditional approach to structural estimation is based on imposing enough restrictions on the structural parameters to achieve point identification, that is, a unique mapping from the reduced-form parameters $\mathbf{\Phi}$ and $\mathbf{\Omega}$ into the structural parameters $\mathbf{A}, \mathbf{B}$, and $\mathbf{D}$. Often these

[2]Specifying $\mathbf{D}$ to be diagonal is motivated by the idea that the structural shocks are primitive and do not have common causes. The idea behind the structural model is to characterize contemporaneous relations between the variables in terms of off-diagonal elements of $\mathbf{A}$; see Bernanke (1986).

come in the form of restrictions that certain elements of $\mathbf{A}$ have to be zero. One of literally thousands of studies that we could use to illustrate this approach is the Cholesky identification scheme used by Kilian (2009). In his application, $\mathbf{y}_t$ is a (3×1) vector consisting of monthly measures of the quantity of oil produced globally, world real economic activity, and the real price of oil: $\mathbf{y}_t = (q_t, y_t, p_t)'$. The number of lags m was taken to be 24, and the system of structural equations can be written[3]

$$q_t = \alpha_{qy} y_t + \alpha_{qp} p_t + \mathbf{b}_1' \mathbf{x}_{t-1} + u_{1t} \tag{6}$$

$$y_t = \alpha_{yq} q_t + \alpha_{yp} p_t + \mathbf{b}_2' \mathbf{x}_{t-1} + u_{2t} \tag{7}$$

$$p_t = \alpha_{pq} q_t + \alpha_{py} y_t + \mathbf{b}_3' \mathbf{x}_{t-1} + u_{3t}. \tag{8}$$

Equation (6) models the behavior of oil producers with α_{qp} the short-run price elasticity of supply. Equation (7) describes the determinants of global economic activity and (8) describes the behavior of oil consumers, with α_{pq} the reciprocal of the short-run price elasticity of demand.

Kilian (2009) achieved identification by imposing $\alpha_{qy} = \alpha_{qp} = 0$, which implies that producers do not respond immediately to changes in economic activity or the price, and $\alpha_{yp} = 0$, meaning that the price of oil has no contemporaneous effect on economic activity. Column 1 of Figure 1 plots the responses of the three variables to a negative shock to the supply of oil under this traditional Cholesky identification scheme.[4] An oil production shortfall is followed a few months later by a modest increase in price and decrease in economic activity.

2.2 A generalization of the traditional approach.

Unfortunately, convincing zero restrictions are a rare commodity in economics. They are hard to come up with and are often difficult to defend. Bayesian methods allow us to incorporate prior information in a much less restrictive way. The Bayesian idea is to represent prior information in the form of a probability distribution over the structural parameters. The value of the prior density $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$ is higher for values of parameters that we think are more plausible a priori and lower for those that we think are less plausible. The value could be zero for parameter configurations that we are absolutely certain can be ruled out. Algorithms for implementing this approach are available at https://sites.google.com/site/cjsbaumeister/research.

Baumeister and Hamilton (2019) showed how the above Cholesky example can be viewed as a special case of the more general Bayesian approach. We used degenerate priors for α_{qy},

[3] The empirical illustrations below are taken from Baumeister and Hamilton (2019), where a detailed description of the data can be found.

[4] Kilian (2009) measured the monthly growth in oil production at an annual rate, whereas we are measuring it at a monthly rate. This is why the units in the (1,1) panel of our Figure 1 are 1/12 the units reported in the (1,1) panel of Kilian's Figure 3.

α_{qp}, and α_{yp} centered at zero and with zero variance to represent the inference of an analyst who was absolutely certain that these three parameters had to be zero. We used priors with fat tails and a huge variance for α_{yq}, α_{pq}, α_{py}, thus treating essentially any value for these parameters as reasonable. The impulse-response functions that result from Bayesian inference using this prior distribution are plotted in the second column of Figure 1. Not surprisingly, this is identical to the first column. The Cholesky approach to identification can be given a Bayesian interpretation, where the prior information is that we claim to know with certainty that the elements in the upper-triangular block of $\mathbf{A}$ were all zero but have no useful information at all about elements in the lower-triangular block of $\mathbf{A}$.

Expressing the prior information that is implicit in the traditional approach in this way invites us to consider a natural generalization. We may have good information that the short-run supply elasticity α_{qp} is small. But it is hard to claim that we know with absolute certainty that α_{qp} has to be zero. On the other hand, we may also have some useful information that the short-run demand elasticity is not large in absolute value. The identifying information that we lose by becoming less dogmatic about the supply elasticity can be compensated in part by incorporating imperfect prior information about other structural parameters like the demand elasticity. Baumeister and Hamilton (2019) illustrated how this idea could be implemented in practice.

2.3 Sign-restricted vector autoregressions.

Another popular approach to structural identification imposes constraints not that parameters like α_{qp} have to be zero but instead that they cannot be negative. Methods for inference based on inequality constraints were developed by Faust (1998), Canova and De Nicoló (2002), Uhlig (2005) and Rubio-Ramírez, Waggoner and Zha (2010) (hereafter RWZ). The RWZ algorithm generates a draw for the reduced-form parameters $\mathbf{\Phi}$ and $\mathbf{\Omega}$ from a Normal-inverted-Wishart distribution and then transforms each $\mathbf{\Omega}$ into a draw for the contemporaneous impacts of one-standard-deviation structural shocks ($\mathbf{H}^*= \mathbf{A}^{-1}\mathbf{D}^{1/2}$) using a QR decomposition of a matrix of standard Normals that comes from the researcher's random-number generator.[5] Draws are only retained if they satisfy the sign and other restrictions that the researcher imposes. The set of retained draws is then typically summarized in terms of 68% intervals around the median.

This approach can also be described as Bayesian inference. Implicit in the draws generated by the RWZ algorithm is a prior probability distribution $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$ over the structural parameters. For example, Baumeister and Hamilton (2015) showed that conditional on $\mathbf{\Omega}$ the RWZ algorithm implies a Cauchy distribution over parameters like α_{qp}. The common prac-

[5] Specifically, the researcher generates an $(n \times n)$ matrix $\mathbf{G}$ consisting of independent standard Normals, factors this as $\mathbf{G} = \mathbf{QR}$ where $\mathbf{Q}$ is an orthonormal matrix and $\mathbf{R}$ is upper triangular, and then generates $\mathbf{H}^*$ from $\mathbf{H}^* = \mathbf{PQ}$ for $\mathbf{P}$ the Cholesky factor of $\mathbf{\Omega}$.

tice of reporting medians and 68% probability regions of the retained draws would be fine if the researcher viewed the implicit prior $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$ as accurately incorporating information that the researcher had about the structural model before seeing the data. However, the typical empirical application makes no defense or even acknowledgement of the implicit prior distribution. If one uses literally no prior information other than the sign restrictions, as many applied studies seem to claim, there is no statistical basis for reporting only 68% of the retained draws. The boundaries of the 68% region are simply an artifact of the researcher's random number generator, not a summary of some feature in the data. If researchers claim not to have used any identifying information other than the sign restrictions, then what they should be reporting is the identified set, namely, *all* the retained values. For more discussion see Watson (2019) and Baumeister and Hamilton (2020a).

2.4 Identification using sign and other restrictions.

Unfortunately, if the identification is based on sign restrictions alone, the identified set is typically too large to be useful. The conclusion that "anything is possible" is not very informative. For this reason, many applied researchers impose additional restrictions along with the sign conditions to throw out more of the draws proposed by the RWZ algorithm; examples include Kilian and Murphy (2012, 2014), Riggi and Venditti (2015), Carter et al. (2017), De Santis and Zimic (2018), and Foroni and Stracca (2019). For illustration we examine the application in Kilian and Murphy (2012). They analyzed the 3-variable system (6)-(8) over almost the same sample as Kilian (2009). They discarded draws that did not satisfy the sign restrictions

$$\operatorname{sign}(\mathbf{H}^*) = \begin{bmatrix} + & + & + \\ + & + & - \\ - & + & + \end{bmatrix}. \tag{9}$$

They further discarded any draw that implied a value for the short-run supply elasticity α_{qp} that was greater than 0.0258.[6] The impulse-response functions that they arrived at are plotted in the third column of Figure 1.[7] Although motivated by a seemingly different identification strategy, the graphs in column 3 are remarkably similar to those in column 1. Restricting the short-run supply elasticity to fall in $(0, 0.0258)$ turns out in practice to be very similar to forcing it to exactly equal zero.

Again if we viewed this algorithm as resulting from a Bayesian prior distribution (which Kilian and Murphy did not), the claim would be that we know on the basis of information before seeing the data that an elasticity of 0.0257 is plausible but a value of 0.0259 is completely

[6] Many other studies using a number of different variables have also imposed a 0.0258 bound on the oil supply elasticity. Examples include Kilian and Murphy (2014), Güntner (2014), Basher et al. (2018), and Geiger and Scharler (2019).

[7] These correspond to the solid lines in row 1 of Figure 5 in Kilian and Murphy (2012).

impossible. A more natural representation of prior information might take the form that while we think a value of 0.0259 is less likely than a value of 0.0257, we cannot say that a value of 0.0259 is completely impossible.

Other researchers have tried to arrive at sharper inference by imposing dynamic sign restrictions. For example, the prior implicit in Kilian and Murphy (2014) maintains that the researcher knows before seeing the data not just the signs of the effect of an oil supply disruption u_{1t} on production, economic activity, and price upon impact ($\partial q_{t+s}/\partial u_{1t}, \partial y_{t+s}/\partial u_{1t}, \partial p_{t+s}/\partial u_{1t}$ for $s = 0$) but also knows with certainty the signs of the effects $s = 1, 2, ..., 12$ months after the shock as well. From the Bayesian perspective, this amounts to a joint prior distribution $p(\mathbf{A}, \mathbf{B})$ over the contemporaneous and lagged structural coefficients. Baumeister and Hamilton (2015) showed how to perform Bayesian analysis when prior information takes this form. However, Canova and Paustian (2011) cautioned that researchers' confidence about the signs of general equilibrium effects at higher horizons is often misplaced. Baumeister and Hamilton (2015, Section 4) argued that we might typically expect effects at horizon s to be similar to those at horizon $s - 1$, with decreasing confidence in that prior information the larger is s. See Baumeister and Hamilton (2018, Table 2) and (2019, page 1896) for illustrations of how this works in practice.

2.5 Relaxing the influence of prior information.

By representing prior information in the form of a density $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$, the analyst is formally summarizing how much we trust the prior information. If we are very certain, the prior should have a small variance, and will end up having a big influence on the posterior distribution. If we are very uncertain, we would use a large variance, and the prior will have little influence on the posterior distribution.

To illustrate alternative uses of prior information, we examine in further detail the implicit prior in Kilian and Murphy (2012, 2014), which maintained that a value for the short-run oil supply elasticity of 0.0257 was plausible but a value of 0.0259 was completely impossible. Where did such precise information come from? Kilian and Murphy (2012) obtained the cutoff 0.0258 from observations in August 1990 after Iraq invaded Kuwait. They divided the increase in oil production in countries other than Iraq and Kuwait in August of 1990 (1.17%) by the increase in oil price in August of 1990 (45.3%), finding $1.17/45.3 = 0.0258$. However, Caldara, Cavallo and Iacoviello (2019) noted that the reason that world production only increased by 1.17% was because of a 19.5% cut from United Arab Emirates, who feared they too would be invaded if they did not cut production. In August 1990, the increase in production from countries other than Iraq, Kuwait and U.A.E. was 1.95%. If one used that figure instead of Kilian and Murphy's 1.17%, the implied elasticity would have been $1.95/45.3 = 0.043$.

Zhou (2020) accordingly used 0.04 as the upper bound on the oil supply elasticity. Again this took the form of an implicit prior belief that a value of 0.039 is plausible while 0.041 is completely impossible. But Caldara, Cavallo and Iacoviello identified dozens of other episodes like August 1990 in which a few countries' production was affected by specific disruptions. Examples include strikes by Norwegian oil workers, Libyan civil conflict, and hurricanes in the Gulf of Mexico. In each of these episodes we can measure the price change and increase of production outside of the impacted countries. The authors used all episodes like this together as instrumental variables to estimate the short-run world oil supply elasticity. They arrived at an estimate of 0.081 with a standard error of 0.037. In a different analysis of individual wells in North Dakota, Bjørnland, Nordvik and Rohrer (2019) concluded, "while output from conventional wells appear non-responsive to price fluctuations in the short-term, we find supply elasticity to be positive and in the range of 0.3-0.9 for shale oil wells, depending on wells and firms characteristics."

Given this prior evidence, we suggest that researchers would not want to impose a strict upper bound, but instead should use a continuous distribution that attaches less and less probability to larger and larger values to summarize existing knowledge. One option is the distribution in Baumeister and Hamilton (2019), which used a Student t with location parameter 0.1, scale parameter 0.2, 3 degrees of freedom, and truncated to be positive.[8] This distribution has a standard deviation of 0.25, reflecting confidence that the parameter is likely well below unity, but not restricting it much beyond that. Our overall results were broadly consistent with those in Kilian and Murphy (2012, 2014), though we concluded that supply shocks had somewhat larger effects than they did.

By contrast, the uniform $(0, 0.0258)$ prior that is implicit in Kilian and Murphy (2012, 2014) has a standard deviation of only 0.0074. This very small variance and dogmatic upper bound of the prior causes it to have a big influence on the results, and explains the remarkable correspondence between columns 1 and 3 of Figure 1.[9]

As a robustness check, Baumeister and Hamilton (2019) also used a prior that assigns an 80% probability that the supply elasticity is uniform over the Kilian-Murphy range of $(0, 0.0258)$ and only a 20% weight on the truncated Student t(0.1,0.2,3). This mixture prior resulted in the identical conclusions as our baseline analysis.[10]

[8] Contrary to the assertions in Kilian and Zhou (2019, 2020), Baumeister and Hamilton's algorithm for Bayesian inference can use any bounded density for $p(\mathbf{A})$. There are no computational or other considerations forcing us to use a truncated Student t distribution. We often choose Student t because it includes the Cauchy distribution implicit in the RWZ algorithm as a special case (a Cauchy variable is a Student t with one degree of freedom) as well as including the Normal distribution as a special case (a Normal variable is a Student t with infinite degrees of freedom). Using 3 degrees of freedom gives the prior much fatter tails than the Normal but still ensures that the prior has a finite variance, which can be an advantage for certain objectives such as calculating a posterior mean.

[9] Kilian and Zhou (2020, p. 22) criticized the Bayesian approach, arguing, "the problem is that, in practice, the existing literature provides little guidance about the nature of this prior." This criticism is misdirected. The analyses in Kilian and Murphy (2012, 2014) and Zhou (2020) make much stronger use of prior information than anything in Baumeister and Hamilton (2019).

[10] Compare column 1 with column 3 of Table 3 and row 1 with row 4 in Table 4 in Baumeister and Hamilton

To summarize, the Bayesian approach allows the researcher to put as much or as little reliance on prior information as is warranted, including both the extremes that one knows some features with certainty ($p(\mathbf{A}, \mathbf{B}, \mathbf{D}) = 0$ for some values) or that one knows essentially nothing about other features (a prior that is flat or nearly flat over broad ranges). Many researchers have missed this fundamental insight that Bayesian inference is a strict generalization of conventional approaches, somehow thinking that prior information is something that is only being used in applications like Baumeister and Hamilton (2019) but not in the thousands of studies that attempt to do structural inference in vector autoregressions without openly acknowledging the Bayesian interpretation of what they are doing.[11]

3 Objects of interest in structural inference.

One difference between the parameterization used by Baumeister and Hamilton (2015, 2019) and that used in most sign-restricted VARs is that Baumeister and Hamilton represent the structural model in terms of the parameters $(\mathbf{A}, \mathbf{B}, \mathbf{D})$ whereas most sign-restricted VARs have focused on the matrix of contemporaneous impacts of one-standard-deviation structural shocks, $\mathbf{H}^* = \mathbf{A}^{-1}\mathbf{D}^{1/2}$. 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. However, if the goal is to use prior information as a tool to allow us to draw structural conclusions from the observed correlations – and we have argued this in fact is the core challenge for structural VARs – this information typically comes in the form of information about $\mathbf{A}$ rather than $\mathbf{H}^*$. Most microfounded models take the form of a system like (1), in which individual equations represent the actions of different agents such as consumers, firms, or government, rather than in the form of postulated general equilibrium impacts of the actions of individual agents. Formulating a prior $p(\mathbf{A}, \mathbf{B}, \mathbf{D})$ typically involves looking at previous findings about elasticities (Baumeister and Hamilton, 2019; Aastveit, Bjørnland, and Cross, 2020), policy rules (Baumeister and Hamilton, 2018; Belongia and Ireland, forthcoming), and responses of agents to permanent changes (Baumeister and Hamilton, 2015). All of these are most naturally represented as information about $\mathbf{A}$,

(2019). We also obtained very similar results for a prior that puts a weight of 95% on the Kilian and Murphy prior.

[11] For example, Kilian and Lütkepohl (2017, p. 455) questioned "why a prior that merely reflects the personal views of the user should be of wider interest to other economists." This misses the key point: the conclusions of *every* structural VAR reflect the analysts' prior information. Indeed, we view the quoted statement as a very strong argument *in support* of our recommendation to communicate carefully the basis for the prior information and to place less weight on those findings that are more in doubt. On page 457 they argued that examining previous studies cannot tell us "what the functional form of the prior density should be. Nor does it tell us what the dispersion of the prior density should be." The dispersion and fatness of the tails of the prior density summarize not the conclusions of previous research but rather the credibility of those conclusions and how broadly they are accepted. Less reliable information is captured by using a prior density with more dispersion, which gives the prior less influence on the posterior inference.

not $\mathbf{A}^{-1}$.

Moreover, many applications of sign-restricted VARs that use the $\mathbf{H}^*$ parameterization have also tried to draw conclusions about behavioral elasticities. Examples include Kilian and Murphy (2012, 2014), Güntner (2014), Riggi and Venditti (2015), Kilian and Lütkepohl (2017), Ludvigson, Ma, and Ng (2017), Antolín-Díaz and Rubio-Ramírez (2018), Basher et al. (2018), Herrera and Rangaraju (2020), and Zhou (2020). These researchers based their estimates of elasticities on the ratios of certain elements of $\mathbf{H}^*$. However, such calculations do not in fact calculate behavioral elasticities, as we now explain.

The issue can be illustrated using a 3-equation system similar to that studied in Section 2.1 with the lagged terms dropped and the demand equation reparameterized. The observed variables are the log of quantity (q_t), the log of income (y_t), and the log of price (p_t),

$$q_t = \gamma y_t + \alpha p_t + u_t^s \tag{10}$$

$$y_t = \xi q_t + \psi p_t + u_t^y \tag{11}$$

$$q_t = \delta y_t + \beta p_t + u_t^d. \tag{12}$$

These correspond to the supply equation, the income equation, and the demand equation, respectively. The parameter β (the demand price elasticity) is the answer to the question: if the price were to increase by 1% with income held constant, by how much would the quantity demanded by consumers change?

How can we estimate this magnitude? Suppose for example that we knew the values of $\gamma, \alpha, \xi, \psi$ and that the structural shocks were mutually uncorrelated. Then the supply shock $u_t^s = q_t - \gamma y_t - \alpha p_t$ and income shock $u_t^y = y_t - \xi q_t - \psi p_t$ would be valid instruments for purposes of estimating the parameters of the demand equation, because they are correlated with y_t and p_t but uncorrelated with u_t^d:

$$\begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T u_t^s y_t & \sum_{t=1}^T u_t^s p_t \\ \sum_{t=1}^T u_t^y y_t & \sum_{t=1}^T u_t^y p_t \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T u_t^s q_t \\ \sum_{t=1}^T u_t^y q_t \end{bmatrix}. \tag{13}$$

Alternatively, we could estimate δ and β by maximizing the likelihood function conditional on u_t^s and u_t^y. For this example, the conditional MLE is numerically identical to the IV estimates; see the appendix for mathematical details.

More generally, using the likelihood function of the observed vector of data $\mathbf{y}_t$ along with any additional information about structural parameters is the optimal way to form inference about the structural parameters. As noted by Rothenberg (1973), from a frequentist perspective, when the additional information takes the form of complete identifying assumptions, maximum likelihood estimation is optimal in the sense that it achieves the smallest asymptotic variance. From a Bayesian perspective, when the additional information takes the form of

Bayesian prior distributions, Bayesian inference is optimal in the sense of minimizing posterior expected loss. Thus the unified approach to incorporating prior information that we are advocating is the optimal way to form an inference about parameters like β, the price elasticity of demand.

Now consider the relation between the parameter β and the observed impacts of structural shocks. The above structural model can be written $\mathbf{A}\mathbf{y}_t = \mathbf{u}_t$, with the impacts of the structural shocks on the observed variables captured by the matrix

$$\begin{aligned} \mathbf{H} &= \frac{\partial \mathbf{y}_t}{\partial \mathbf{u}_t'} = \mathbf{A}^{-1} \\ &= |\mathbf{A}|^{-1} \begin{bmatrix} -\beta - \delta\psi & \alpha\delta - \beta\gamma & \alpha + \gamma\psi \\ -\psi - \beta\xi & \alpha - \beta & \psi + \alpha\xi \\ \delta\xi - 1 & \delta - \gamma & 1 - \gamma\xi \end{bmatrix}. \end{aligned} \tag{14}$$

What would we get if we tried to estimate the demand elasticity on the basis of the ratio of the change in q_t to the change in p_t in response to a shock to supply u_t^s? For this system that ratio is given by[12]

$$\frac{h_{11}}{h_{31}} = \frac{-\beta - \delta\psi}{\delta\xi - 1}. \tag{15}$$

In general, expression (15) is not the demand elasticity β. The reason is that if there is a shock to u_t^s, not only will it change the price p_t, but it will also change income. The size of the change in price is $|\mathbf{A}|^{-1}(\delta\xi - 1)$ and the size of the change in income is $|\mathbf{A}|^{-1}(-\psi - \beta\xi)$. From the demand curve, the change in price will lead to a change in quantity demanded of β times the change in price, namely $\beta|\mathbf{A}|^{-1}(\delta\xi - 1)$. Likewise the change in income will lead to a change in quantity demanded of δ times the change in income, namely $\delta|\mathbf{A}|^{-1}(-\psi - \beta\xi)$. The observed change in quantity demanded in response to the shock in supply is the sum of these two terms,

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

Dividing this by the magnitude of the change in price that results from the supply shock, $|\mathbf{A}|^{-1}(\delta\xi - 1)$, produces the result (15).

In the special case when demand does not respond to income ($\delta = 0$), expression (15) would simplify to the correct answer β. But in general, expression (15) reflects a combination of the sensitivity of demand to price and the sensitivity of demand to income.

For example, when Kilian and Murphy (2014) calculated the demand elasticity in their 4-variable model they used an incorrect expression like (15). To calculate the supply elasticity they used two different expressions, both of which are incorrect. They first calculated h_{12}/h_{32},

[12] This is identical to the ratio h_{11}^*/h_{31}^* for $\mathbf{H}^* = \mathbf{H}\mathbf{D}^{1/2}$.

the ratio of the first to the third elements of the second column of $\mathbf{A}^{-1}$. This is the ratio of the change in quantity to the change in price in response to what they call a flow demand shock. They next calculated h_{13}/h_{33}, the ratio of the first and third elements of the third column of $\mathbf{A}^{-1}$. This is the ratio of the change in quantity to the change in price in response to what they call a speculative 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. The way their code works is to reject a draw unless both of these proposed measures of the supply elasticity are below 0.0258.

In the correct structural interpretation, there is a single unique magnitude that should be called the supply elasticity. This is the (1,3) element of $\mathbf{A} = \mathbf{H}^{-1}$, and this is the object for which optimal inference is obtained by following the unified Bayesian approach.[13]

Notwithstanding, the researcher 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 $\mathbf{H}$. There is no problem incorporating information about $\mathbf{H}$ as a supplement to information about $\mathbf{A}$. Suppose that for the system given by (10)-(12) (and the necessary implication of those three equations in the form of expression (14)) we had prior information about both the price elasticity of supply $p_1(\alpha)$ and the contemporary 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 $\mathbf{A}$. 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) and Grisse (2020) all incorporate prior information about both $\mathbf{A}$ and $\mathbf{A}^{-1}$.[14]

[13] Kilian and Zhou (2020) have separately argued that one cannot talk about the elasticity of oil demand at all in a 3-variable system like (6)-(8) because some purchased oil is not consumed as fuel but is instead stored as inventory by refiners. This argument is fallacious. The demand price elasticity in the 3-variable system summarizes the price sensitivity of the joint response of consumers and refiners which is a perfectly well defined and economically relevant concept. Baumeister and Hamilton (2019) in fact analyzed a 4-variable system that includes inventories, and Baumeister and Hamilton (2020b) demonstrated that those are correctly measuring the concept in which Kilian and Zhou (2020) claim we should be interested, whereas the calculations based on $\mathbf{H}^*$ in Kilian and Murphy (2014) do not.

[14] 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." Their statement is false. The method described by Baumeister and Hamilton (2018, 2019) for incorporating information about both $\mathbf{A}$ and $\mathbf{A}^{-1}$ is completely general.

4 Computational considerations.

Discarding draws from the RWZ algorithm that do not satisfy specified conditions is vulnerable to the concern of arbitrariness that we highlighted above. We think it is unreasonable to treat a value $\alpha_{qp} = 0.0257$ as perfectly plausible but rule out $\alpha_{qp} = 0.0259$ as completely impossible. An additional consideration is that as the criteria for rejecting draws becomes more strict, the number of accepted draws shrinks. For example, the code for Kilian and Murphy (2014) that is 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 have been discussing up to this point, which is whether the identifying assumptions being applied (in this case, that $\alpha_{qp} < 0.0258$) are really credible. Second is the practical issue of what to conclude from the 16 retained draws. If the goal is to rely on no prior information beyond the sign and other restrictions explicitly imposed, we argued above that researchers should be reporting the identified set, that is, the set of all values consistent with the observed data and the imposed restrictions. Sixteen observations is not enough to estimate the full extent of this set. The fact that we end up with only a couple of usable realizations from every million draws suggests at a minimum that the RWZ algorithm for models like this is so inefficient that we don't reliably find the answer we're looking for using the method as typically implemented.

Moreover, researchers typically report not the entire set of retained draws but instead just highlight a summary statistic such as the median retained draw. For example, Kilian and Murphy (2014, p. 464) reported results based on the "model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity among the admissible structural models." 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 2. 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, and reproduce 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 2 show what these imputed effects look like 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 enables us to characterize the Bayesian posterior distribution for any objects of interest quite accurately in a way that will be identical when one makes a change as trivial as changing the random number seed. Moreover, Bayesian statistical decision theory spells out exactly how we should use the posterior distribution to summarize the findings. For detailed discussion of loss functions and optimal statistical inference see Baumeister and Hamilton (2018).

5 Estimating structural vector autoregressions using instrumental variables or proxy variables.

Another promising approach to identification in structural VARs has been proposed by Stock and Watson (2012, 2018) and Mertens and Ravn (2014). The idea is to find a proxy or instrumental variable that is correlated with one of the structural shocks of interest. For example, Känzig (2019) proposed to use the market response in a narrow window around OPEC announcements as an instrument for the supply shock. If this variable is uncorrelated with the other structural shocks, then it is a valid instrument and provides an alternative way to achieve identification. Noh (2019) and Paul (forthcoming) showed that such IV estimates can be easily obtained using restricted OLS regressions.

Stock and Watson (2018) noted that the validity of the instrument in conjunction with the auxiliary assumptions that are typically implicit in these applications is empirically testable. Noh (2019) and Paul (forthcoming) developed a particularly simple way to conduct this test. Unfortunately, often this test rejects the hypothesis that instruments are valid. Even if it does not reject, many researchers might still have some doubts about the instruments.

Nguyen (2019) suggested that we could treat potential instruments or proxies in a similar way to other identifying information. The identifying assumption that z_t is a valid instrument for structural shock u_{it} is a claim that the population correlation between z_t and u_{jt} for $j \neq i$ is exactly zero. A strict generalization of this would be a prior belief that the correlation is close to zero, though we are not 100% certain of this. Nguyen showed how the Bayesian approach can be used to combine imperfect confidence in the instruments with uncertain prior information about other aspects of the structure. This allows us to perform inference in a system in which we openly acknowledge doubts about both the validity of instruments and about our other identifying information and to incorporate these doubts into statistical summaries of what we can conclude from the observed data.

6 Conclusions.

Many researchers treat identifying assumptions as a necessary evil, seeing the task to be to make enough of them to get sharp answers to questions of interest. We suggest that researchers instead begin by thinking carefully about the meaning of the structural parameters, looking for information that we can obtain about these from other datasets, model calibrations, or economic theory. The next step is to summarize that knowledge in the form of a prior density $p(\mathbf{A})$ that accurately reflects both the information and our doubts, for example, using large variances for features we honestly know little about. Once prior information has been represented in this way, it's simply a matter of plugging their subroutine to calculate $p(\mathbf{A})$ into the code posted at https://sites.google.com/site/cjsbaumeister/research.

Identification is not an either-or decision of whether the researcher should use hard restrictions, sign restrictions, or proxy variables. They can all be used together, even if we have doubts about each one of them. A Bayesian perspective that unifies all the different approaches into a common framework makes it possible to take advantage of the strengths of each method while incorporating honest doubts about each into the final statistical summary.

References

Aastveit, Knut Are, Hilde C. Bjørnland, and Jamie L. Cross (2020). “Inflation Expectations and the Pass-through of Oil Prices,” working paper, Norwegian Business School.

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

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 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.

Baumeister, Christiane, and James D. Hamilton (2020a). “Drawing Conclusions from Structural Vector Autoregressions Identified on the Basis of Sign Restrictions,” NBER working paper 26606.

Baumeister, Christiane, and James D. Hamilton (2020b). “Structural Interpretation of Vector Autoregressions with Incomplete Identification: Setting the Record Straight,” working paper, UCSD.

Belongia, Michael T., and Peter N. Ireland (forthcoming). “A Classical View of the Business Cycle,” *Journal of Money, Credit and Banking.*

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

Bjørnland, Hilde C., Frode Martin Nordvik, and Maximilian Rohrer (2019). “Supply Flexibility in the Shale Patch: Evidence from North Dakota,” working paper, Norwegian Business School.

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

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

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

Carter, Colin A., Gordon C. Rausser, and Aaron Smith (2017). “Commodity Storage and the Market Effects of Biofuel Policies,” *American Journal of Agricultural Economics* 99(4):

1027-1055.

De Santis, Roberto A., and Srečko Zimic (2018). "Spillovers Among Sovereign Debt Markets: Identification Through Absolute Magnitude Restrictions," *Journal of Applied Econometrics* 33: 727–747.

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

Foroni, Claudia, and Livio Stracca (2019). "Much Ado About Nothing? The Shale Oil Revolution and the Global Supply Curve," ECB working paper 2309.

Geiger, Martin, and Johann Scharler (2019). "How Do Consumers Assess the Macroeconomic Effects of Oil Price Fluctuations? Evidence from U.S. Survey Data," *Journal of Macroeconomics* 62: 1-19.

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.

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.

Känzig, Diego R. (2019). "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements," working paper, London Business School.

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 (2019). "Oil Supply Shock Redux?" Working paper.

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

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

Mertens, Karel, and Morten O. Ravn (2014). "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers," *Journal of Monetary Economics* 68: S1-S19.

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

Noh, Eul (2019). "Impulse-response Analysis with Proxy Variables," working paper, SSRN.

Paul, Pascal (forthcoming). "The Time-Varying Effect of Monetary Policy on Asset Prices," *Review of Economics and Statistics*.

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.

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.

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: The relation between IV and MLE in a simple 3-variable example.

Equation (13) characterized the instrumental-variables estimates of elasticities in a simple 3-variable example, which we used to explain how economists typically approach estimation of elasticities in simultaneous-equation systems. IV is closely related to (and in general, is less efficient than) full-information maximum likelihood, which is the frequentist analog to Bayesian inference about structural parameters using the likelihood function. In this appendix we show that, for this simple 3-variable example, IV turns out to be identical to maximum likelihood. Our purpose in doing so is to help explain why analysis based on the likelihood function implied by the structural model (whether frequentist or Bayesian) is the correct and indeed the optimal way to estimate elasticities.

Characterization of IV estimate.

Expression (13) can be rewritten

$$\begin{bmatrix} \sum_{t=1}^T u_t^s y_t & \sum_{t=1}^T u_t^s p_t \\ \sum_{t=1}^T u_t^y y_t & \sum_{t=1}^T u_t^y p_t \end{bmatrix} \begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T u_t^s q_t \\ \sum_{t=1}^T u_t^y q_t \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T q_t y_t & \sum_{t=1}^T q_t p_t \\ \sum_{t=1}^T y_t y_t & \sum_{t=1}^T y_t p_t \\ \sum_{t=1}^T p_t y_t & \sum_{t=1}^T p_t p_t \end{bmatrix} \begin{bmatrix} \hat{\delta}_{IV} \\ \hat{\beta}_{IV} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T q_t q_t \\ \sum_{t=1}^T y_t q_t \\ \sum_{t=1}^T p_t q_t \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T q_t q_t & \sum_{t=1}^T q_t y_t & \sum_{t=1}^T q_t p_t \\ \sum_{t=1}^T y_t q_t & \sum_{t=1}^T y_t y_t & \sum_{t=1}^T y_t p_t \\ \sum_{t=1}^T p_t q_t & \sum_{t=1}^T p_t y_t & \sum_{t=1}^T p_t p_t \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{\delta}_{IV} \\ -\hat{\beta}_{IV} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{16}$$

We can partition $\mathbf{A}$ as

$$\underset{(3\times3)}{\mathbf{A}} = \begin{bmatrix} \underset{(2\times3)}{\mathbf{\Gamma}} \\ \underset{(1\times3)}{\boldsymbol{\eta}'} \end{bmatrix}$$

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

$$\boldsymbol{\eta}' = \begin{bmatrix} 1 & -\delta & -\beta \end{bmatrix}.$$

This allows (16) to be written compactly as

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

for $\hat{\mathbf{\Omega}}$ the sample variance-covariance matrix of the observed data:

$$\hat{\mathbf{\Omega}} = \begin{bmatrix} T^{-1}\sum_{t=1}^{T} q_t q_t & T^{-1}\sum_{t=1}^{T} q_t y_t & T^{-1}\sum_{t=1}^{T} q_t p_t \\ T^{-1}\sum_{t=1}^{T} y_t q_t & T^{-1}\sum_{t=1}^{T} y_t y_t & T^{-1}\sum_{t=1}^{T} y_t p_t \\ T^{-1}\sum_{t=1}^{T} p_t q_t & T^{-1}\sum_{t=1}^{T} p_t y_t & T^{-1}\sum_{t=1}^{T} p_t p_t \end{bmatrix}.$$

Characterization of MLE.

Next consider maximum likelihood estimation. Conditional on the (2×1) vector $\mathbf{z}_t = \mathbf{\Gamma}\mathbf{y}_t$, the remaining randomness in $\mathbf{y}_t$ can be summarized in terms of the observed scalar $w_t = \boldsymbol{\gamma}_\perp' \mathbf{y}_t$ where $\boldsymbol{\gamma}_\perp$ is the (3×1) vector that is orthogonal to the rows of $\mathbf{\Gamma}$ and whose third element is normalized to be unity.[15] If $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{D})$, then $\mathbf{y}_t$ is multivariate Normal and $w_t | \mathbf{z}_t \sim N(\boldsymbol{\pi}(\delta, \beta)' \mathbf{z}_t, v)$ where given $\mathbf{\Gamma}$, the (2×1) vector $\boldsymbol{\pi}$ is a known function of (δ, β). To calculate this function, notice that $\mathbf{y}_t = \mathbf{A}^{-1}\mathbf{u}_t$ so $w_t = \boldsymbol{\gamma}_\perp' \mathbf{y}_t = \boldsymbol{\gamma}_\perp' \mathbf{A}^{-1}\mathbf{u}_t$. Since the elements of $\mathbf{u}_t$ are mutually independent, the coefficients in the population projection of w_t on (u_t^s, u_t^y) are given by the first two terms of the vector $\boldsymbol{\gamma}_\perp' \mathbf{A}^{-1}$:

$$\boldsymbol{\pi}(\delta, \beta)' = \boldsymbol{\gamma}_\perp' \mathbf{A}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{18}$$

Given $\mathbf{\Gamma}$, this is a known function of (δ, β).

The conditional MLE of (δ, β) is the value that maximizes

$$-(T/2)\log(2\pi) - (T/2)\log v - \frac{\sum_{t=1}^{T}[w_t - \boldsymbol{\pi}(\delta, \beta)' \mathbf{z}_t]^2}{2v}$$

or equivalently the value that minimizes $\sum_{t=1}^{T}[w_t - \boldsymbol{\pi}(\delta, \beta)' \mathbf{z}_t]^2$. Note that in general the value of $\boldsymbol{\pi}$ that minimizes this sum of squared residuals is the OLS estimate

$$\hat{\boldsymbol{\pi}}_{OLS}' = \left(\sum_{t=1}^{T} w_t \mathbf{z}_t'\right)\left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'\right)^{-1}. \tag{19}$$

[15] That is,

$$\begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\gamma}_\perp = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{\gamma}_\perp = \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note that $\boldsymbol{\gamma}_\perp$ is known from $\mathbf{\Gamma}$ and does not depend on (δ, β).

We claim that if we choose $(\hat{\delta}_{MLE}, \hat{\beta}_{MLE})$ such that

$$\hat{\boldsymbol{\eta}}'_{MLE} \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{z}'_t\right) \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}'_t\right)^{-1} = \mathbf{0}', \tag{20}$$

then (18) will satisfied:

$$\hat{\boldsymbol{\pi}}'_{OLS} = \boldsymbol{\gamma}'_{\perp} \begin{bmatrix} 1 & -\gamma & -\alpha \\ -\xi & 1 & -\psi \\ 1 & -\hat{\delta}_{MLE} & -\hat{\beta}_{MLE} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{21}$$

To verify (21), first define the (3×2) matrix

$$\begin{aligned} \hat{\boldsymbol{\Pi}} &= \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{z}'_t\right) \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}'_t\right)^{-1} \\ &= \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}'_t \boldsymbol{\Gamma}'\right) \left(\sum_{t=1}^{T} \boldsymbol{\Gamma} \mathbf{y}_t \mathbf{y}'_t \boldsymbol{\Gamma}'\right)^{-1} \end{aligned} \tag{22}$$

so that for example

$$\hat{\boldsymbol{\pi}}'_{OLS} = \boldsymbol{\gamma}'_{\perp} \hat{\boldsymbol{\Pi}}. \tag{23}$$

Premultiply (22) by $\hat{\mathbf{A}}_{MLE}$:

$$\begin{bmatrix} \boldsymbol{\Gamma} \\ \hat{\boldsymbol{\eta}}'_{MLE} \end{bmatrix} \hat{\boldsymbol{\Pi}} = \begin{bmatrix} \boldsymbol{\Gamma} \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}'_t \boldsymbol{\Gamma}'\right) \left(\sum_{t=1}^{T} \boldsymbol{\Gamma} \mathbf{y}_t \mathbf{y}'_t \boldsymbol{\Gamma}'\right)^{-1} \\ \hat{\boldsymbol{\eta}}'_{MLE} \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}'_t \boldsymbol{\Gamma}'\right) \left(\sum_{t=1}^{T} \boldsymbol{\Gamma} \mathbf{y}_t \mathbf{y}'_t \boldsymbol{\Gamma}'\right)^{-1} \end{bmatrix}. \tag{24}$$

The first two rows of (24) will be recognized as the (2×2) identity matrix, and the third row is zero by the proposed choice for $\hat{\boldsymbol{\eta}}_{MLE}$. Thus

$$\hat{\mathbf{A}}_{MLE} \hat{\boldsymbol{\Pi}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Premultiplying by $\boldsymbol{\gamma}'_{\perp} \hat{\mathbf{A}}_{MLE}^{-1}$ gives

$$\boldsymbol{\gamma}'_{\perp} \hat{\boldsymbol{\Pi}} = \boldsymbol{\gamma}'_{\perp} \hat{\mathbf{A}}_{MLE}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

or using (23),

$$\hat{\boldsymbol{\pi}}'_{OLS} = \boldsymbol{\gamma}'_{\perp} \hat{\mathbf{A}}_{MLE}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

as claimed in (21).

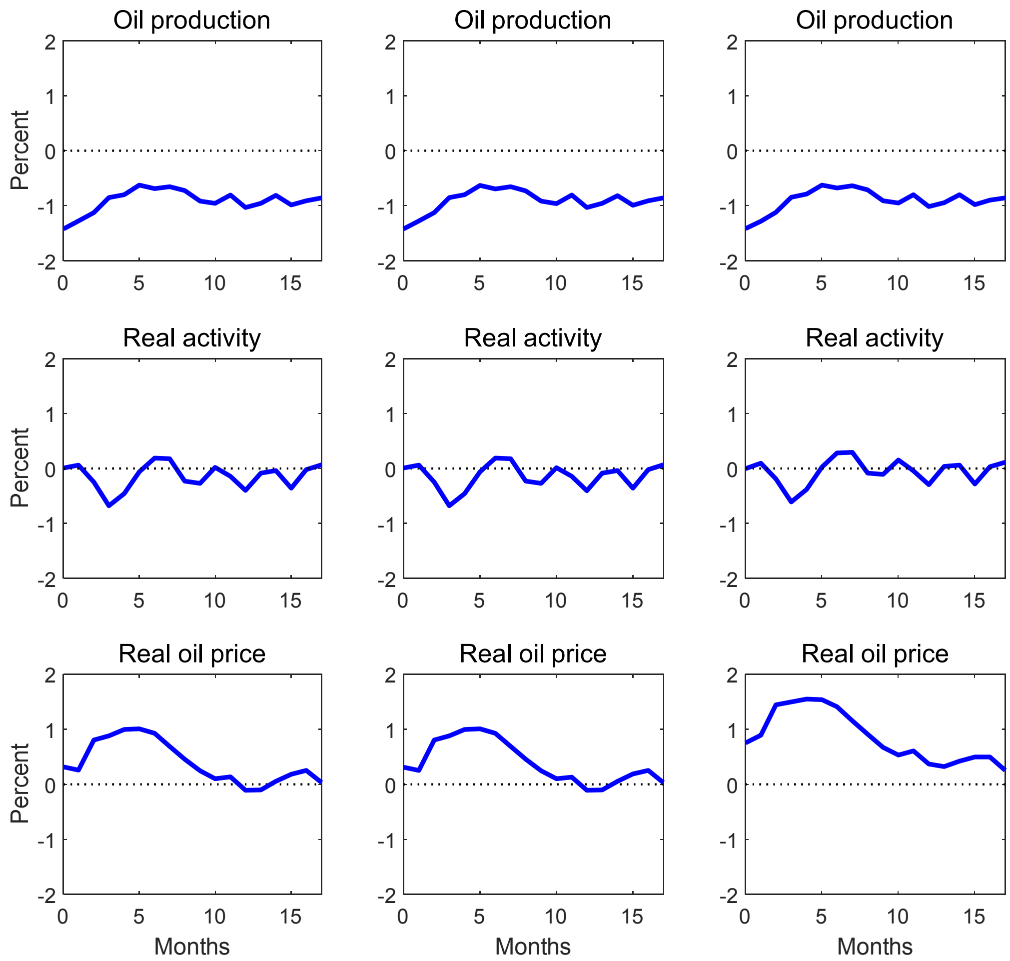
Demonstration of equivalence.

Note that since $\left(\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'\right)$ has full rank, condition (20) can equivalently be written

$$\begin{aligned}\mathbf{0}' &= \hat{\boldsymbol{\eta}}_{MLE}'\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{z}_t'\right)\\ &= \hat{\boldsymbol{\eta}}_{MLE}'\left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right)\boldsymbol{\Gamma}'.\end{aligned}$$

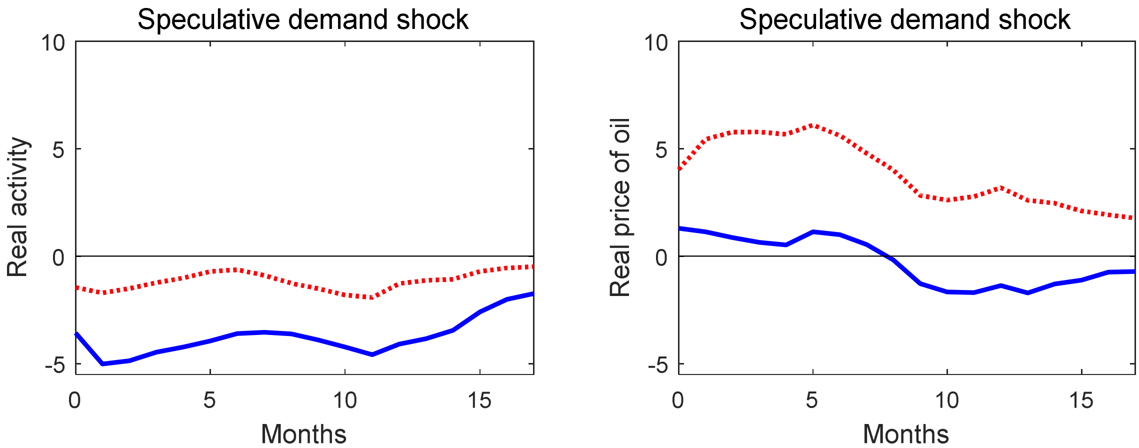
This is simply the transpose of (17), confirming that for this example MLE is numerically identical to IV.

Figure 1. Impulse-response functions for the effects of an oil supply shock under three characterizations of the identifying assumptions.



Notes to Figure 1. Effects of a one-standard-deviation decrease in u_{1t} on $\boldsymbol{y}_{t+s}$ using the Choleski identification as in Kilian (2009) (column 1), using the Bayesian prior described in the text (column 2), and using the identification scheme in Kilian and Murphy (2012) (column 3).

Figure 2. 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.



Notes to Figure 2. 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.