Structural Vector Autoregressions with Imperfect Identifying Information

By Christiane Baumeister and James D. Hamilton*

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.

In this paper we propose a unifying principle for dealing with uncertainty about identifying assumptions. 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. 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 by increasing the variance. 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.

I. Identification in structural vector autoregressions

A vector autoregression is a convenient way to summarize the dynamic correlations that we observe in the data. For $\mathbf{y}_t$ a vector of n different variables observed at date t, we can regress each variable on a constant and m lags of all of the variables:

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

We can use the regression estimates to forecast the variables in the system. But to draw causal conclusions, we would need a structural model such as

$$(2) \quad \mathbf{A}\mathbf{y}_t = \mathbf{k} + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{B}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{B}_m \mathbf{y}_{t-m} + \mathbf{u}_t.$$

For example, suppose that $\mathbf{y}_t$ contains observations on 3 variables: measures of global oil production, world economic activity, and the real price of oil. We might view these variables as determined by 3 structural equations. The first row of equation (2) models the behavior of oil producers, with the (1,3) element of $\mathbf{A}$ determined by the short-run price elasticity of oil supply. The second row of (2) describes the determinants of economic activity and the third row the behavior of oil consumers. If we knew the values of the coefficients in (2) we could make statements about the dynamic effects of a disruption in oil supply, that is, the consequences over time of a shock to the first element of $\mathbf{u}_t$.

A. *The traditional approach to identification*

The traditional approach to identification is to place enough restrictions on the structural parameters in (2) to be able to estimate the remaining structural parameters directly

* Baumeister: Department of Economics, University of Notre Dame (email: cjsbaumeister@gmail.com). Hamilton: Department of Economics, University of California at San Diego (email: jhamilton@ucsd.edu).

from the OLS regression estimates in (1). One common approach is to assume a recursive form to the structural model in which the upper-triangular elements of $\mathbf{A}$ are all zero and the structural shocks $\mathbf{u}_t$ are uncorrelated with each other and across time. In this case the remaining structural parameters can be easily estimated from the regression estimates and a Cholesky factorization of the variance matrix of the forecast errors in (1). This approach to identification amounts to assuming, for example, that oil producers respond to changes in economic activity or oil prices only with a lag, meaning that the short-run price elasticity of oil supply is zero. An example of this approach is the analysis in Kilian (2009). The first column of Figure 1 reproduces his structural estimates of the dynamic effects on oil production, economic activity, and the oil price of a one-standard-deviation decrease in u_1, along with 68% and 95% confidence intervals. These results suggest that a decrease in oil production is followed by a decrease in economic activity and increase in oil prices.

B. A Bayesian interpretation of the traditional approach

By contrast, the Bayesian approach to structural inference begins with a prior probability distribution that characterizes our beliefs about the structural parameters before seeing the data. This is represented for example in the form of a density $p(\mathbf{A})$ that assigns higher probabilities to values of $\mathbf{A}$ that are more plausible on the basis of economic theory or analysis of other data sets and lower probabilities to values that are less plausible. The density could be zero for values that we are certain can be ruled out altogether. The Bayesian then updates these distributions to take into account how that understanding is changed by the observed data. Many economists dislike the idea of bringing in prior beliefs about the structure and allowing those beliefs to influence our conclusions. But we would argue that the traditional approach to identification is doing exactly this. In Baumeister and Hamilton (2019) we applied the algorithm developed for Bayesian inference in Baumeister and Hamilton (2015) to the dataset in Kilian (2009) with a particular prior for $\mathbf{A}$ in which the diagonal elements of $\mathbf{A}$ are normalized at unity. We used degenerate priors for the upper-triangular elements 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 the lower-triangular elements, thus treating essentially any value for these parameters as reasonable. The results of that Bayesian inference are reported 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. The Bayesian prior that is implicit in Cholesky identification claims to know with certainty that the upper-triangular elements of $\mathbf{A}$ are all zero but has no useful information at all about the lower-triangular elements.

C. A Bayesian generalization of the traditional approach

Framing the traditional approach to identification as prior certainty about some aspects of the structure invites us to consider a natural generalization – what if we have doubts about the identifying assumptions themselves? One natural way to represent this in the previous example is to use a prior density for the upper-triangular elements of $\mathbf{A}$ that has most of its mass near zero but does not insist these parameters have to exactly equal zero. For illustration we use a Student t distribution with scale parameter of 0.05 and 3 degrees of freedom. The small scale parameter means that we think it unlikely that the value would be larger than 0.1 in absolute value. A Student t has fatter tails than the Normal, which means we would be slightly less surprised by values significantly outside of ± 0.1 than if we used a Normal distribution. The results are shown in the third column of Figure 1. The point estimates are similar to those for Cholesky identification, but the posterior credibility bands in this case are wider.

Unlike the first two columns, the third model is unidentified in the formal econo-

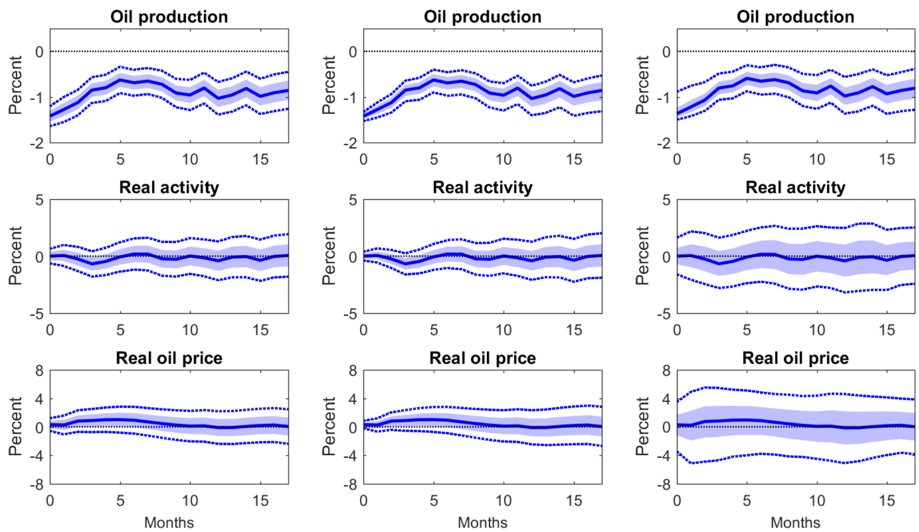

FIGURE 1. EFFECTS OF DECREASE IN OIL SUPPLY UNDER DIFFERENT IDENTIFICATION SCHEMES. COLUMN 1: CONVENTIONAL CHOLESKY. COLUMN 2: BAYESIAN INTERPRETATION OF CHOLESKY. COLUMN 3: BAYESIAN WITH LESS CONFIDENCE IN CHOLESKY RESTRICTIONS.

metric sense. The error bands in the first two columns reflect uncertainty that results because we do not know the true values of the reduced-form parameters in (1) perfectly but had to estimate them from a limited dataset. If we had more observations, that uncertainty would become smaller, and would eventually vanish with an infinite sample size. By contrast, the bands in the third column represent both uncertainty about reduced-form parameters and uncertainty about how we give those parameters a structural interpretation. The first source of uncertainty would disappear as the sample size gets larger, but the second would not. In general we would have a set of possible values for the structural parameters, known as the identified set, that would remain as possible answers even if the sample size was infinite. Baumeister and Hamilton (2015) characterized the identified set and noted that the Bayesian approach allows us to rank observationally equivalent outcomes based on their prior plausibility. Baumeister and Hamilton (2018) showed if we want to minimize the absolute error of our inference, the medians of the posterior distribution (represented by the solid blue lines in the second and third columns of Figure 1) provide the optimal estimate of the dynamic structural effects given both the finite data set and our imperfect understanding of the structure.

This Bayesian interpretation invites another natural question – are the three upper-triangular elements of **A** the only aspects of the economic structure about which we have any prior information? For many applied researchers, these zeros were imposed as a matter of necessity – I need 3 zeros for identification, and well, here are three. In the example here, we do have good information that the short-run price elasticity of supply is indeed small. But we also have good information that the short-run price elasticity of demand is small, which tells us something about the (3,1) element of **A**. We also know from economic theory that supply should respond positively to a price increase and demand respond negatively, and know something about plausible effects of oil prices on economic activity. By bringing in information from a variety of sources – all of it imperfect, none of it telling us the value of any structural parameter with certainty – we can offset some of the loss in accuracy that we surrendered when we acknowledged doubts about the zeros that were required for traditional identification. Baumeister and Hamilton (2019) illustrated how this can be done in an analysis of the oil market, while Baumeister and Hamilton (2018) examined the effects of monetary policy.

When the prior information about supply and demand elasticities comes from different sources, it is natural to model the joint prior as the product of the separate priors. We may also have prior information about how the different elements of $\mathbf{A}$ interact. For example, Baumeister and Derdzyan (2021) argued that the price elasticity of gasoline demand and crude oil demand are fundamentally related. Baumeister and Hamilton (2018, 2019) showed how to incorporate prior information about aspects of $\mathbf{A}^{-1}$, which summarizes how the elements of $\mathbf{A}$ interact to determine the equilibrium impacts of structural shocks.

D. Identification using sign restrictions

A number of researchers have relied on prior information about the signs of equilibrium impacts. For example, we might expect a negative oil supply shock to lead to a shortfall in oil production, a slowdown of economic activity, and a rise in oil prices, which would imply negative values for the (1,1) and (2,1) elements and a positive value for the (3,1) element of $\mathbf{A}^{-1}$. A convenient algorithm for performing inference imposing such restrictions was developed by Uhlig (2005) and Rubio-Ramirez, Waggoner and Zha (2010). This algorithm generates possible answers drawn from a certain distribution and retains the draws if they satisfy the sign restrictions. Most users of this approach want to report point estimates and error bands like those in Figure 1, yet do not think that they have used any information other than the sign restrictions in arriving at their estimates. However, all answers within the identified set satisfy the sign restrictions and are observationally equivalent. If a researcher has used no identifying information other than the sign restrictions, there is no basis for selecting one of the draws generated by the algorithm (such as the median) as a representative estimate. Moreover, researchers should be reporting *all* the draws, not just a subset containing 68% or 95% of the draws.

One could give a Bayesian interpretation to this approach if the distribution from which possible answers are drawn truly reflects prior information about the *structural* model. If that were the case, this prior information would provide a ranking of possible outcomes in the identified set and a basis for reporting the median draw as the optimal point estimate; see Baumeister and Hamilton (2015). However, in the typical sign-restriction study this ranking is implicit in the algorithm and is not controlled by the researcher nor influenced by economic theory. Our recommendation is that if researchers want to report point estimates, they need to explain the source of the prior information that allows them to rank observationally equivalent outcomes.

The algorithms that we have developed allow researchers to do exactly this. The asymmetric t distribution developed by Baumeister and Hamilton (2018) allows one to incorporate prior information that the sign of a specified element of $\mathbf{A}^{-1}$ is quite likely to be positive or negative, without being absolutely certain, and do so separately from information about the likely magnitude.

E. Identification using instrumental variables

Stock and Watson (2012) and Mertens and Ravn (2013) developed another promising approach to identification that uses instrumental variables or proxies to identify structural shocks. Potential instruments must be correlated with the structural shock of interest but uncorrelated with the other structural shocks. Again researchers often have doubts about this assumption in practical applications. A strict generalization of the usual assumption behind the IV approach is the prior belief that the correlation with other shocks is close to zero, though we are not 100% certain of this. Nguyen (2019) 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.

II. 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

Baumeister, Christiane, and Aram Derdzyan. 2021. "Structural Changes in Energy Markets." Unpublished.

Baumeister, Christiane, and James D Hamilton. 2015. "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information." *Econometrica*, 83(5): 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 Identification: Revisiting the Role of Oil Supply and Demand Shocks." *American Economic Review*, 109(5): 1873–1910.

Kilian, Lutz. 2009. "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market." *American Economic Review*, 99(3): 1053–69.

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(4): 1212–47.

Nguyen, Lam. 2019. "Bayesian Inference in Structural Vector Autoregression with Sign Restrictions and External Instruments." https://sites.google.com/view/nguyenlam/research.

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

Stock, James H, and Mark W Watson. 2012. "Disentangling the Channels of the 2007-2009 Recession." *Brookings Papers on Economic Activity*, 43(1): 81–156.

Uhlig, Harald. 2005. "What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure." *Journal of Monetary Economics*, 52(2): 381–419.