**Abstract:** Structural Vector Autoregressions (SVARs) are a multivariate, linear representation of a vector of observables on its own lags. SVARs are used by economists to recover economic shocks from observables by imposing a minimum of assumptions compatible with a large class of models. This article reviews, first, the relation of SVARs with dynamic stochastic general equilibrium models. Second, it discusses the normalization, identification, and estimation of SVARs. The article finishes with an assessment of the advantages and drawbacks of SVARs.

# Structural Vector Autoregressions

Structural Vector Autoregressions (SVARs hereafter) are a multivariate, linear representation of a vector of observables on its own lags and (possibly) other variables as a trend or a constant. SVARs make explicit identifying assumptions to isolate estimates of policy and/or private agents' behavior and its effects on the economy while keeping the model free of the many additional restrictive assumptions needed to give every parameter a behavioral interpretation. Introduced by Sims (1980), SVARs have been used to document the effects of money on output (Sims and Zha, 2005), the relative importance of supply and demand shocks on business cycles (Blanchard and Quah, 1989), the effects of fiscal policy (Blanchard and Perotti, 2002), or the relation between technology shocks and worked hours (Galí, 1999), among many other applications.

## 1. Economic theory and the SVAR representation

Dynamic economic models can be viewed as restrictions on stochastic processes. Under this perspective, an economic theory is a mapping between a vector of $k$ economic shocks $w_t$ and a vector of $n$ observables $y_t$ of the form $y_t = D\left(w^t\right)$, where $w^t$ represents the whole history of shocks $w_t$ up to period $t$. The economic shocks are those shocks to the fundamental elements of the theory: preferences, technology, informational sets, government policy, measurement errors, etc. The observables are all variables that the researcher has access to. Often, $y_t$

includes a constant to capture the mean of the process. The mapping $D(\cdot)$ is the product of the equilibrium behavior of the agents in the model, implied by their optimal decision rules and consistency conditions like resource constraints and market clearing. The construction of the mapping $D(\cdot)$ is the sense in which economic theory tightly relates shocks and observables. Also, the mapping $D(\cdot)$ can be interpreted as the impulse response of the model to an economic shock.

Often, we restrict our attention to linear mappings of the form $y_t = D(L) w_t$, where $L$ is the lag operator. For simplicity of exposition, $w_t$ will be *i.i.d.* random variables and normally distributed, $w_t \sim N(0, \Sigma)$. More involved structures, e.g., allowing for autocorrelation between the shocks, can be accommodated with additional notation.

We pick the neoclassical growth model, the workhorse of dynamic macroeconomics, to illustrate the previous paragraphs. In its basic version, the model maps productivity shocks, the $w_t$ of the theory, into observables, $y_t$, like output or investment. The mapping comes from the optimal investment and labor supply decisions of the households, the resource constraint of the economy, and the law of motion for productivity. If the productivity shocks are normally distributed and we solve the model by linearizing its equilibrium conditions, we obtain a mapping of the form $y_t = D(L) w_t$ described above.

If $k = n$, i.e., we have as many economic shocks as observables, and $|D(L)|$ has all its roots outside the unit circle, we can invert the mapping $D(L)$ (see Fernández-Villaverde, Rubio-Ramírez, and Sargent, 2005) and obtain:

$$A(L) y_t = w_t$$

where $A(L) = A_0 - \Sigma_{k=1}^{\infty} A_k L^k$ is a one-sided matrix lag polynomial that embodies all the (usually non-linear) cross-equation restrictions derived by the equilibrium solution of the model. In general, $A(L)$ is of infinite order. This representation is known as the SVAR representation. The name comes from realizing that $A(L) y_t = w_t$ is a Vector Autoregression (VAR) generated by an economic model (a "structure").

## 2. Reduced–form representation, normalization, and identification

Consider now the case where a researcher does not have access to the SVAR representation. Instead, she has access to the VAR representation of $y_t$:

$$y_t = B_1 y_{t-1} + B_2 y_{t-2} + ... + a_t,$$

where $E y_{t-j} a_t = 0$ for all $j$ and $E a_t a_t' = \Omega$. This representation is known as the reduced-form representation. Can the researcher recover the SVAR representation using the reduced-form representation? Fernández-Villaverde, Rubio-Ramírez, and Sargent (2005) show that, given a strictly invertible economic model, i.e., $|D(L)|$ has all its roots strictly outside the unit circle, there is one and only one identification scheme to recover the SVAR from the reduced form. In addition, they show that the mapping between $a_t$ and $w_t$ is $a_t = A_0^{-1} w_t$. Hence, if the researcher knew $A_0^{-1}$, she could recover the SVAR representation from the reduced-form, noting that $A_j = A_0 B_j$ for all $j$ and $w_t = A_0 a_t$.

Hence, the recovery of $w_t$ from $y^T$ requires the knowledge of the dynamic economic model. Can we avoid this step? Unfortunately, the answer is, in general, no, because knowledge of the reduced-form matrices $B_i$'s and $\Omega$ does not imply, by itself, knowledge of the $A_i$'s and $\Sigma$. There are two reasons for that.

The first is normalization. Reversing the signs of two rows or columns of the $A_i$'s does not matter for the $B_i$'s. Thus, without the correct normalization restrictions, statistical inference about the $A_i$'s is essentially meaningless. Waggoner and Zha (2003) provide a general normalization rule that maintains coherent economic interpretations.

The second is identification. If we knew $A_0$, each equation $B_i = A_0^{-1} A_i$ would determine $A_i$ given some $B_i$. But the only restrictions that the reduced-form representation imposes on the matrix $A_0$ comes from $\Omega = A_0 \Sigma A_0'$. In this relationship, we have $n(3n+1)/2$ unknowns (the $n^2$ distinct elements of $A_0$ and the $n(n+1)/2$ distinct elements of $\Sigma$) for $n^2$ knowns (the $n(n+1)/2$ distinct elements of $\Omega$). Thus, we require $n^2$ identification restrictions. Since we can set the diagonal elements of $A_0$ equal to 1 by scaling, we are left with the need of

$n(n-1)$ additional identification restrictions (alternatively, we could scale the shocks such that the diagonal of $\Sigma$ is composed of ones and leave the diagonal of $A_0$ unrestricted). These identification restrictions are dictated by the economic theory being studied.

The literature, however, has often preferred to impose identification restrictions that are motivated by the desire to be compatible with a large class of models, instead of just one concrete model and its whole set of cross-equations restrictions. The hope is that, thanks to this generality, the inferences drawn from SVAR can be more robust and can compensate for the lack of efficiency derived from not implementing a full information method.

The most common identification restriction has been to assume that $\Sigma$ is diagonal. This assumption relies on the view that economic shocks are inherently different sources of uncertainty that interact only through their effect on the decisions of the model's agents. Since this assumption imposes $n(n-1)/2$ restrictions, we still require $n(n-1)/2$ additional restrictions.

To find these additional restrictions, economists have followed two main approaches: short-run restrictions and long-run restrictions. Sims (1980) pioneered the first approach when he proposed to impose zeroes on $A_0$. The motivation for such a scheme comes from the idea that there is a natural timing in the effect of economic shocks. For example, to place a zero on $A_0$, we can use the intuition that monetary policy cannot respond contemporaneously to a shock in the price level because of informational delays. Similarly, institutional constraints, like the timing of tax collections, can be exploited for identification (Blanchard and Perotti, 2002). Sims (1980) ordered variables in such a way that $A_0$ is lower triangular. Sims and Zha (2005) present a non-triangular identification scheme on an eight-variable SVAR.

The long-run restrictions were popularized by Blanchard and Quah (1989). These restrictions are imposed on $A(1) = A_0 - \Sigma_{k=1}^{\infty} A_k$. Since $A^{-1}(1) = D(1)$, long-run restrictions are justified as restrictions on the long-run effects of economic shocks, usually on the first difference of an observable. For example, Blanchard and Quah (1989) assume that there are two shocks ("demand" and "supply") affecting unemployment and output. The demand shock has no long-run effect on unemployment or output. The supply shock has no long-run effect on unemployment but may have a long-run effect on output. These differences in their long-run impacts allow Blanchard and Quah to identify the shocks and trace their impulse response function.

New identification schemes have been proposed to overcome the difficulties of the existing approaches. See, for instance, Uhlig (2005) for an identification scheme of monetary policy shocks based on sign restrictions that hold across a large class of models.

## 3. Estimation

Why is the previous discussion of the relation between the reduced and structural form of a VAR relevant? Because the reduced form can be easily estimated. An empirically implementable version of the reduced-form representation truncates the number of lags at the $p$-th order:

$$y_t = \overline{B}_1 y_{t-1} + ... + \overline{B}_p y_{t-p} + \overline{a}_t$$

where $E\overline{a}_t\overline{a}_t' = \overline{\Omega}$. We use hats in the matrices and the error $\overline{a}_t$ to indicate that they do not correspond exactly to the reduced form of the model but to the truncated version. The effects of the truncation on the accuracy of inference delivered by SVARs are unclear (see Chari, Kehoe, McGrattan, 2005, and Christiano, Eichenbaum, and Vigfusson, 2005, for two opposite assessments). The resulting truncated VAR can be taken to the data using standard methods: GMM, Maximum Likelihood, or Bayesian.

The Bayesian approach is especially attractive. SVARs are proliferatively parameterized. The number of parameters in $B(L)$ grows with the square of the number of variables and the number of lags. Consequently, given the short period of data typically available to macroeconomists, classical methods become unreliable. A careful use of prior information alleviates the problem of overparameterization and improves the quality of the inference. The advent of modern simulation techniques, especially Markov Chain Monte Carlo methods, have made the implementation of the Bayesian paradigm straightforward, even for sophisticated priors.

The point estimates $\widehat{\widehat{B}}_i$ for $i = 1, .., p$ and $\widehat{\widehat{\Omega}}$ can be used to find estimates of $\widehat{A}_i$ and $\widehat{\Sigma}$ by solving $\widehat{\widehat{B}}_i = \widehat{A}_0^{-1}\widehat{A}_i$ for $i = 1, ..., p$, and $\widehat{\widehat{\Omega}} = \widehat{A}_0\widehat{\Sigma}\widehat{A}_0'$. With an estimate of $\widehat{A}_0$ and the $\overline{a}_t$, we can get $\widehat{w}_t = \widehat{A}_0\widehat{\overline{a}}_t$. Thus, the reduced form plus the identifying restrictions deliver both an estimate of the economic shocks as well as the impulse response of the variables in the economy to those shocks. Confidence intervals for point estimates and error bands for impulse response functions can be estimated resorting to Markov Chain Monte Carlo techniques or the bootstrap.

In an important recent contribution, Sims and Zha (2006) have extended the estimation of SVARs to allow for changes in equation coefficients and variances. This paper opens the door for the analysis of richer dynamic models with parameter instability, arguably a more realistic description of observed aggregate variables.

## 4. Assessment of SVARs

SVARs offer an attractive approach to estimation. They promise to coax interesting patterns from the data that will prevail across a set of incompletely specified dynamic economic models with a minimum of identifying assumptions. Moreover, SVARs are easy to estimate, being possible to do so even with commercial software and freely available routines from the Internet. In the hands of skillful researchers, SVARs have contributed to the understanding of aggregate fluctuations, have clarified the importance of different economic shocks, and have generated fruitful debates among macroeconomists.

However, SVARs have also been criticized. We mention only three criticisms. First, it has been argued that the economic shocks recovered from an SVAR do not resemble the shocks measured by other mechanisms, such as market expectations embodied in future prices. Second, the shocks recovered from an SVAR may reflect variables omitted from the model. If these omitted variables correlate with the included variables, the estimated economic shocks will be biased. Third, the results of many SVAR exercises, even simple ones, are sensitive to the identification restrictions. Related to this criticism is the view that many of the identification schemes are the product of a specification search in which researchers look for "reasonable" answers. If an identification scheme matches the conventional wisdom, it is called successful; if it does not is called a puzzle or, even worse, a failure (Uhlig, 2005). Consequently, there is a danger that economists will get stuck in an priori view of the data under the cloak of formal statistical inference.

Jesús Fernández-Villaverde

*University of Pennsylvania, NBER, and CEPR*

and

Juan F. Rubio-Ramírez

*Federal Reserve Bank of Atlanta*

## References


[1] Blanchard, O.J. and D. Quah 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances. American Economic Review 79, 655-673.

[2] Blanchard, O.J. and R. Perotti 2002. An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. Quarterly Journal of Economics 117, 1329-1368.

[3] Chari, V.V., P.J. Kehoe, and E.R. McGrattan 2005. A Critique of Structural VARs Using Real Business Cycle Theory. Federal Reserve Bank of Minneapolis Working Paper 631.

[4] Christiano, L.J., M. Eichenbaum, and R. Vigfusson 2005. Assessing Structural VARs. Mimeo, Northwestern University.

[5] Fernández-Villaverde, J., J.F. Rubio-Ramírez, and T.J. Sargent 2005. A,B,C's (and D's) for Understanding VARs. NBER Technical Working Paper 308.

[6] Galí, J. 1999. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? American Economic Review 89, 249-271.

[7] Sims, C.A. 1980. Macroeconomics and Reality. Econometrica 48, 1-48.

[8] Sims, C.A. and T. Zha 2005. Does Monetary Policy Generate Recessions? Macroeconomic Dynamics, forthcoming.

[9] Sims, C.A. and T. Zha 2006. Were There Regime Switches in U.S. Monetary Policy? American Economic Review, forthcoming.

[10] Uhlig, H. 2005. What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. Journal of Monetary Economics 52, 381-419.

[11] Waggoner, D.F., and T. Zha 2003. Likelihood Preserving Normalization in Multiple Equation Models. Journal of Econometrics 114, 329-347.