On the Identification of Structural Vector Autoregressions

Pierre-Daniel G. Sarte

Following seminal work by Sims (1980a, 1980b), the economics profession has become increasingly concerned with studying sources of economic fluctuations. Sims's use of vector autoregressions (VARs) made it possible to address both the relative importance and the dynamic effect of various shocks on macroeconomic variables. This type of empirical analysis has had at least two important consequences. First, by deepening policymakers' understanding of how economic variables respond to demand versus supply shocks, it has enabled them to better respond to a constantly changing environment. Second, VARs have become especially useful in guiding macroeconomists towards building structural models that are more consistent with the data.

According to Sims (1980b), VARs simply represented an atheoretical technique for describing how a set of historical data was generated by random innovations in the variables of interest. This reduced-form interpretation of VARs, however, was strongly criticized by Cooley and Leroy (1985), as well as by Bernanke (1986). At the heart of the critique lies the observation that VAR results cannot be interpreted independently of a more structural macroeconomic model. Recovering the structural parameters from an estimation procedure requires that some restrictions be imposed. These are known as identifying restrictions. Implicitly, the choice of variable ordering in a reduced-form VAR constitutes such an identifying restriction.

As a result of the Cooley-Leroy/Bernanke critique, economists began to focus more precisely upon the issue of identifying restrictions. The extent to which specific innovations were allowed to affect some subset of variables,

I would like to thank Tom Cooley, Michael Dotsey, Bruce Hansen, Tom Humphrey, Yash Mehra, and Alex Wolman for more than helpful comments. I would also like to thank Sergio Rebelo, Vassilios Patikis, and Mark Watson for their suggestions. The opinions expressed herein are the author's and do not represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

either in the short run or in the long run, began to be derived explicitly from structural macroeconomic models. Consequently, what were previously considered random surprises could be interpreted in terms of specific shocks, such as technology or fiscal policy shocks. This more refined use of VARs, known as structural vector autoregressions (SVARs), has become a popular tool for evaluating economic models, particularly in the macroeconomics literature.

The fact that nontrivial restrictions must be imposed for SVARs to be identified suggests, at least in principle, that estimation results may be contingent on the choice of restrictions. To take a concrete and recent example, in estimating a system containing employment and productivity variables, Gali (1996) achieves identification by assuming that aggregate demand shocks do not affect productivity in the long run. Using postwar U.S. data, he is then able to show that, surprisingly, employment responds negatively to a positive technology shock. One may wonder, however, whether his results would change significantly under alternative restrictions. This article consequently investigates how the use of different identifying restrictions affects empirical evidence about business fluctuations. Two important conclusions emerge from the analysis.

First, by thinking of SVARs within the framework of instrumental variables estimation, it will become clear that the method is inappropriate for certain identifying restrictions. This finding occurs because SVARs use the estimated residual from a previous equation in the system as an instrument in the current equation. Since estimation of this residual depends on some prior identifying restriction, the identification scheme necessarily determines the strength of the instrument. By drawing from the literature on estimation with weak instruments, this article points out that in some cases, SVARs will not yield meaningful parameter estimates.

The second finding of interest suggests that even in cases where SVAR parameters can be properly estimated, different identification choices can lead to contradictory results. For example, in Gali (1996) the restriction that aggregate demand shocks not affect productivity in the long run also implies that employment responds negatively to a positive technology shock. But the opposite result emerges when aggregate demand shocks are allowed to have a small negative effect on productivity in the long run. This latter restriction is appropriate if demand shocks are interpreted as fiscal policy shocks in a real business cycle model. More importantly, this observation suggests that sensitivity analysis should form an integral part of deciding what constitutes a stylized fact within the confines of SVAR estimation.

This article is organized as follows. We first provide a brief description of reduced-form VARs as well as the basic idea underlying the Cooley-Leroy/Bernanke critique. In doing so, the important assumptions underlying the use of VARs are laid out explicitly for the nonspecialist reader. We then introduce the mechanics of SVARs—that is, the details of how SVARs are usually estimated—and link the issue of identification to the estimation procedure.[1] The next section draws from the literature on instrumental variables in order to show the conditions in which the SVAR methodology fails to yield meaningful parameter estimates. We then describe the type of interpretational ambiguities that may arise when the same SVAR is estimated using alternative identifying restrictions. Finally, we offer a brief summary and some conclusions.

1. REDUCED-FORM VARs AND THE COOLEY-LEROY/BERNANKE CRITIQUE

In this section, we briefly describe the VAR approach first advocated by Sims (1980a, 1980b). In doing so, we will show that the issue of identification already emerges in interpreting estimated dynamic responses for a given set of variables. To make matters more concrete, the analysis in both this and the next section is framed within the context of a generic bivariate system. However, the basic issues under consideration are invariant with respect to the size of the system. Thus, consider the joint time series behavior of the vector $(\Delta y_t, \Delta x_t)$, which we summarize as

$$B(L)Y_t = e_t, \text{ with } B(0) = B_0 = I, \tag{1}$$

where $Y_t = (\Delta y_t, \Delta x_t)'$, and $B(L)$ denotes a matrix polynomial in the lag operator L. $B(L)$ is thus defined as $B_0 + B_1 L + \ldots + B_k L^k + \ldots$, where $L^k Y_t = Y_{t-k}$. Since $B(0) = I$, equation (1) is an unrestricted VAR representation of the joint dynamic behavior of the vector Y_t. In Sims's (1980a) original notation, the vector $e_t = (e_{yt},\ e_{xt})'$ would carry the meaning of "surprises" or innovations in Δy_t and Δx_t respectively.

In its simplest interpretation, the reduced form in (1) is a model that describes how the historical data contained in Y_t was generated by some random mechanism. As such, few would question its usefulness as a forecasting tool. However, in the analysis of the variables' dynamic responses to the various innovations, the implications of the unrestricted VAR are not unambiguous. Specifically, let us rewrite (1) as a moving average representation,

$$Y_t = B(L)^{-1} e_t = C(L) e_t, \tag{2}$$

where $C(L)$ is defined to be equal to $B(L)^{-1}$, with $C(L) = C_0 + C_1 L\ +\ \ldots\ +$ $C_K L^K + \ldots$, and $C_0 = C(0) = B(0)^{-1} = I$. To obtain the comparative dynamic responses of Δy_t and Δx_t, Sims (1980a) first suggested orthogonalizing the vector of innovations e_t by defining $f_t = Ae_t$, such that A is a lower triangular matrix with 1s on its diagonal and f_t has a normalized diagonal covariance

[1] Note that the details of the estimation procedure described in this article apply directly to the work of King and Watson (1997).

matrix. This particular transformation is known as a Choleski factorization and the newly defined innovations, $f_t = (f_{yt}, f_{xt})'$, have unit variance and are orthogonal. Equation (2) can therefore also be expressed as

$$Y_t = C(L)A^{-1}Ae_t = D(L)f_t, \tag{3}$$

with $D(L) = C_0A^{-1} + C_1A^{-1}L + \ldots + C_kA^{-1}L^k + \ldots$. Responses to innovations at different horizons, also known as impulse responses, are then given by

$$E_t\frac{\partial Y_{t+k}}{\partial f_t} = C_kA^{-1}, \text{ for } k = 0, 1, \ldots. \tag{4}$$

The advantage of computing dynamic responses in this way is that the innovations f_t are uncorrelated. Therefore it is very simple to compute the variances associated with any linear combinations involving them. Note that

$$E_{t+1}Y_{t+k} - E_tY_{t+k} = C_kA^{-1}f_t, \tag{5}$$

so that the j^{th} row of C_kA^{-1} gives the marginal effect of f_t on the j^{th} variable's k step-ahead forecast error. Since the f_t's are uncorrelated with unit variance, squaring the elements of C_kA^{-1} leads to contributions of the elements of f_t to the variance of the k step-ahead forecast error. This latter process is known as variance decomposition and describes the degree to which a particular innovation contributes to observed fluctuations in Y_t. Note that the variance decomposition of the contemporaneous forecast error is given by the squared elements of $C_0A^{-1} = A^{-1}$. More importantly, since A is a lower triangular matrix, A^{-1} is also lower triangular. This implies that the innovation in the first equation, f_{yt}, explains 100 percent of the variance in the contemporaneous forecast error of Δy_t. But this is precisely an identifying restriction on the dynamic behavior of Y_t. In a larger system, the variance of the contemporaneous forecast error in the j^{th} variable would be entirely accounted for by the first j innovations in a recursive fashion. Each of these restrictions would then implicitly constitute prior identifying restrictions. In this sense, the ordering of variables in a reduced-form VAR is of crucial significance.

This last point was made, perhaps most vigorously, in Cooley and Leroy (1985): "if the models (i.e., VARs) are interpreted as non-structural, we view the conclusions as unsupportable, being structural in nature. If the models are interpreted as structural, on the other hand, the restrictions on error distributions adopted in atheoretical macroeconometrics are not arbitrary renormalizations, but prior identifying restrictions." On a related note, Bernanke (1986) also writes that the standard Choleski decomposition, while "sometimes treated as neutral . . . in fact embodies strong assumptions about the underlying economic structure." Following these criticisms, several authors, including Blanchard and Watson (1984), Sims (1986), Bernanke (1986), and Blanchard and Quah (1989), addressed the issue of identification explicitly. The error terms in these latter models were given structural interpretations and the results no longer had to depend on an arbitrary orthogonalization. However, this latter methodology possesses its own problems, both in terms of the validity of the estimation procedure and the interpretation of the results. This is the subject to which we now turn our attention.

2. INTRODUCTION TO THE MECHANICS OF STRUCTURAL VARS

The reduced form in equation (1) could simply be thought of as a way to summarize the full data set Y_t. In contrast, suppose that a theoretical model tells us that y_t actually evolves according to a specific stochastic process,

$$\Delta y_t = \Theta_{ya}(L)\epsilon_{at} + \Theta_{yb}(L)\epsilon_{bt} + (1-L)\Phi_{ya}(L)\epsilon_{at} + (1-L)\Phi_{yb}(L)\epsilon_{bt}, \tag{6}$$

where ϵ_{at} and ϵ_{bt} now possess well-defined structural interpretations. Thus, y_t might represent national output, while ϵ_{at} and ϵ_{bt} might denote shocks to technology and labor supply respectively. This specification for y_t is quite general in that it allows shocks to have both permanent and temporary effects. The polynomial in the lag operator $\Phi(L)$ captures temporary deviations in y_t, while the polynomial $\Theta(L)$ keeps track of permanent changes in its steady-state level. Similarly, suppose that x_t follows a process that can be described by

$$\Delta x_t = \Theta_{xa}(L)\epsilon_{at} + \Theta_{xb}(L)\epsilon_{bt} + (1-L)\Phi_{xa}(L)\epsilon_{at} + (1-L)\Phi_{xb}(L)\epsilon_{bt}. \tag{7}$$

With this specification in hand, it is possible to summarize the system as

$$Y_t = S(L)\epsilon_t, \tag{8}$$

where Y_t is defined as in the previous section, $\epsilon_t = (\epsilon_{at}, \epsilon_{bt})'$, and

$$S(L) = \begin{bmatrix} \Theta_{ya}(L) + (1-L)\Phi_{ya}(L) & \Theta_{yb}(L) + (1-L)\Phi_{yb}(L) \\ \Theta_{xa}(L) + (1-L)\Phi_{xa}(L) & \Theta_{xb}(L) + (1-L)\Phi_{xb}(L) \end{bmatrix}. \tag{9}$$

Equation (8) therefore denotes the structural moving average representation of the variables y_t and x_t, as a function of the exogenous innovations ϵ_{at} and ϵ_{bt}. Let us assume that $S(L)$ is invertible so that equation (8) can also be expressed in autoregressive form:

$$T(L)Y_t = S(L)^{-1}Y_t = \epsilon_t, \tag{10}$$

that is,

$$T(L)\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} \epsilon_{at} \\ \epsilon_{bt} \end{bmatrix}, \text{ with } T(0) = S(0)^{-1} \neq I. \tag{11}$$

Since the two exogenous processes that govern the behavior of y_t and x_t in (6) and (7) are assumed stationary, we also assume that the roots of the polynomial matrix $|T(z)|$ lie outside the unit circle. At this stage it is not possible to disentangle the structural effects of ϵ_{at} and ϵ_{bt} in equation (11). Put another way, we cannot currently identify the structural error terms ϵ_{at} and ϵ_{bt} with the residuals in the two equations implicit in (11). This is a well-known problem that naturally leads us to the issue of identification.

Identification in Structural VARs

To get a handle on the problem of identification, observe the relationship between the reduced form in (1) and equation (11). Since $T(L)Y_t = T_0 Y_t + T_1 Y_{t-1} + \ldots$, it follows that $T_0^{-1}T(L)Y_t = Y_t + T_0^{-1}T_1 Y_{t-1} + \ldots = T_0^{-1}\epsilon_t$. We then see that $T_0^{-1}T(L)Y_t$ is the reduced form, that is, $T_0^{-1}T(L) = B(L)$ so that

$$T(0)^{-1}T(L)Y_t = B(L)Y_t = e_t = T(0)^{-1}\epsilon_t. \tag{12}$$

Hence, if $\Sigma = \text{cov}\ (\epsilon_t)$ and $\Omega = \text{cov}\ (e_t)$, the following relation also holds:

$$T(0)^{-1}\Sigma T(0)^{-1\prime} = \Omega. \tag{13}$$

Since Ω can be estimated from the reduced form, the problem of identification relates to the conditions under which the structural parameters in $T(0)^{-1}\Sigma T(0)^{-1\prime}$ can be recovered from Ω. Equation (13) potentially establishes a set of three equations in seven unknowns. Specifically, the unknowns consist of four parameters in $T(0)$ and two variances and one covariance term in Σ. The SVAR literature typically reduces the size of this problem by making the following two assumptions. First, $T(0)$ is normalized to contain 1s on its diagonal. Second, Σ is diagonalized, which reflects the assumption that the structural disturbance terms are taken to be uncorrelated. This leaves us with four unknowns; therefore, one further restriction must be imposed for the structural form to be identified. This additional restriction will generally reflect the econometrician's beliefs and, as will be apparent below, will allow one to separate the effects of the two structural error terms.

As we have just pointed out, only one restriction needs to be imposed upon the dynamics of the system in (11) for the parameters to be identified. One possibility is to specify a priori one of the parameters in the contemporaneous matrix $T(0)$. Another popular approach, the one we focus on here, is to pre-specify a particular long-run relationship between the variables and therefore constrain the matrix of long-run multipliers $T(1)$. This approach is the one followed by Shapiro and Watson (1988), Blanchard and Quah (1989), King, Plosser, Stock, and Watson (1991), and Gali (1992, 1996) among others. To be concrete, define

$$T(1) = \begin{bmatrix} 1-\theta_{yy} & -\theta_{yx} \\ -\theta_{xy} & 1-\theta_{xx} \end{bmatrix} = \begin{bmatrix} \Theta_{ya}(1) & \Theta_{yb}(1) \\ \Theta_{xa}(1) & \Theta_{xb}(1) \end{bmatrix}^{-1} = S(1)^{-1}. \tag{14}$$

One way to achieve identification would be to impose the restriction that the exogenous process with innovation ϵ_{at} not affect the level of x_t in the long run.

That is, impose the restriction that

$$\Theta_{xa}(1) = 0. \tag{15}$$

Since inverses of block diagonal matrices are themselves block diagonal, setting $\Theta_{xa}(1) = 0$ is tantamount to setting $\theta_{xy} = 0$. It would then be possible to estimate all the remaining parameters in equation (6) and (7). This type of restriction, known as an exclusion restriction, is used for identification in the papers cited above. Note, however, that in theory there is no reason why identified parameters should be set to zero as opposed to any other value. All that is required is that the set of identified parameters be fixed in advance, whether zero or not. For example, if ϵ_{at} denotes a shock to technology and x_t represents labor supply, imposing $\Theta_{xa}(1) = 0$ would mean the structural model we have in mind implies that changes in technology do not affect labor supply in the long run. However, in a standard real business cycle model, the permanent effect of technology on labor supply depends on whether the income or the substitution effect dominates. This effect in turn depends on whether the elasticity of intertemporal substitution is greater or less than one. Therefore, there is no reason why exclusion restrictions should necessarily be used as an identification strategy.

The fact that $\Theta_{xa}(1)$, or alternatively θ_{xy}, does not have to be set to zero as a way to identify the model means that estimated parameters, and therefore estimated dynamic responses, can vary depending on the identification scheme adopted. This observation carries with it two potential problems. First, different identification schemes might lead to different comparative dynamic responses of the variables. Therefore, in using SVARs to establish stylized facts, some sensitivity analysis appears to be essential. Second, the estimation procedure may fail in a statistical sense for some values of θ_{xy} in the relevant parameter space. Before looking at each of these problems, however, we first need to explain SVAR estimation.

Structural VAR Estimation Procedure

The most popular way of imposing identifying restrictions as part of the estimation procedure in a SVAR is to take an instrumental variables (IV) approach, specifically two-stage least squares. In applying this approach to our bivariate system, we examine a simple case involving one lag. This will help in keeping matters tractable. Thus, the second equation in (11) can be written as

$$\Delta x_t = \beta_{xy0}\Delta y_t + \beta_{xy1}\Delta y_{t-1} + \beta_{xx1}\Delta x_{t-1} + \epsilon_{bt}. \tag{16}$$

To see how the long-run multipliers θ_{xx} and θ_{xy} in $T(1)$ implicitly enter in equation (16), observe that this equation can also be expressed as

$$\Delta x_t - \theta_{xy}\Delta y_t = \gamma_{xy0}\Delta^2 y_t + \theta_{xx}\Delta x_{t-1} + \epsilon_{bt}, \tag{17}$$

where $\Delta^2 y_t$ denotes the second difference in y_t, $\theta_{xx} = \beta_{xx1}$, $\gamma_{xy0} = -\beta_{xy1}$, and $\theta_{xy} = \beta_{xy0} + \beta_{xy1}$.[2] By setting a predetermined value for θ_{xy}, not necessarily zero, the parameters of equation (17) can then be estimated. Since $\Delta^2 y_t$ is correlated with ϵ_{bt}, ordinary least squares estimation is inappropriate, but two-stage least squares can be performed using the set $\mathbf{Z} = \{\Delta x_{t-1}, \Delta y_{t-1}\}$ as instruments. In a similar fashion, the equation for Δy_t can be written as

$$\Delta y_t = \beta_{yy1}\Delta y_{t-1} + \beta_{yx0}\Delta x_t + \beta_{yx1}\Delta x_{t-1} + \epsilon_{at}. \tag{18}$$

Equation (18) can be estimated using the same set of instruments as for (17) plus the estimated residual for ϵ_{bt}.[3] Recall that in order to achieve identification, the structural disturbances were assumed uncorrelated, thereby allowing the use of the estimated residual as an instrument. Furthermore, this residual is the only candidate instrument that remains. Additional lags of the endogenous variables, if relevant, should have been included in the original equations.

The key point to note at this stage is that since the left-hand side of equation (17) varies with θ_{xy}, the parameters as well as the error term in that equation are contingent upon the identification scheme. This raises a question as to the validity of the estimated residual from equation (17) as an instrument. Not only is zero correlation between the structural disturbances necessary, but a high correlation between the instrument and the variable it is instrumenting for is also essential. This point is emphasized by Nelson and Startz (1990). As we shall now see, because the time series behavior of the estimated residual in (17) varies with θ_{xy}, the validity of the estimation procedure in the subsequent equation will be implicitly tied to the choice of identifying restriction.

3. IDENTIFICATION FAILURE IN STRUCTURAL VARs

To gain insight into the problems that may arise in this framework, given the identification strategy adopted, let us rewrite equation (17) as follows:

$$\Delta x_t - \theta_{xy}\Delta y_t = \mathbf{X}\phi + \epsilon_{bt}, \tag{19}$$

where $\mathbf{X} = \{\Delta^2 y_t, \Delta x_{t-1}\}$ and $\phi = (\gamma_{xy0}, \theta_{xx})'$. Then, the two-stage least squares estimator $\widehat{\phi}$ is given by

$$\widehat{\phi} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\Delta x_t - \theta_{xy}\Delta y_t). \tag{20}$$

From equation (20), the parameter estimates in $\widehat{\phi}$ will change as θ_{xy} takes on different values. This is also true of the estimated residual, which we therefore

[2] As an intermediate step, equation (16) can also be expressed as $\Delta x_t = (\beta_{xy0} + \beta_{xy1} - \beta_{xy1})\Delta y_t + \beta_{xy1}\Delta y_{t-1} + \beta_{xx1}\Delta x_{t-1} + \epsilon_{bt}$.

[3] Observe that, analogously to (17), this equation can also be written as $\Delta y_t = \theta_{yy}\Delta y_{t-1} + \theta_{yx}\Delta x_t + \gamma_{yx0}\Delta^2 x_t + \epsilon_{at}$, where $\theta_{yy} = \beta_{yy1}$, $\gamma_{yx0} = -\beta_{yx1}$, and $\theta_{yx} = \beta_{yx0} + \beta_{yx1}$.

denote by $e_{bt}(\theta_{xy})$ to underscore its dependence on the adopted identification strategy. Since this estimated residual can be computed as

$$\begin{aligned} e_{bt}(\theta_{xy}) &= (\Delta x_t - \theta_{xy}\Delta y_t) - \mathbf{X}\widehat{\phi} \\ &= (\Delta x_t - \theta_{xy}\Delta y_t) - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\Delta x_t - \theta_{xy}\Delta y_t), \end{aligned} \tag{21}$$

observe that $\mathbf{Z}'e_{bt}(\theta_{xy}) = e_{bt}(\theta_{xy})'\mathbf{Z} = 0 \ \forall \theta_{xy}$. This last condition summarizes what are sometimes called the normal equations. Now, the second equation to be estimated in (18) can also be expressed as

$$\Delta y_t = \mathbf{Z}\beta + \Delta x_t \beta_{yx0} + \epsilon_{at}, \tag{22}$$

where $\beta = (\beta_{yx1}, \beta_{yy1})'$, and Δx_t is the endogenous variable of interest. Since the relevant set of instruments for the estimation of equation (22) is given by $\{\mathbf{Z}, e_{bt}(\theta_{xy})\}$, it follows that the two-stage least squares estimator for β is given by

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\beta_{yx0}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\Delta x_t \\ e_{bt}(\theta_{xy})'\mathbf{Z} & e_{bt}(\theta_{xy})'\Delta x_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Z}'\Delta y_t \\ e_{bt}(\theta_{xy})'\Delta y_t \end{bmatrix}. \tag{23}$$

This last expression can be thought of as a set of two equations in two unknowns, specifically,

$$\mathbf{Z}'\mathbf{Z}\widehat{\beta} + \mathbf{Z}'\Delta x_t \widehat{\beta_{yx0}} = \mathbf{Z}'\Delta y_t \tag{24}$$

and

$$e_{bt}(\theta_{xy})'\mathbf{Z}\widehat{\beta} + e_{bt}(\theta_{xy})'\Delta x_t \widehat{\beta_{yx0}} = e_{bt}(\theta_{xy})'\Delta y_t. \tag{25}$$

Therefore it follows that

$$\widehat{\beta_{yx0}} = [e_{bt}(\theta_{xy})'\mathbf{M}_z \Delta x_t]^{-1}[e_{bt}(\theta_{xy})'\mathbf{M}_z \Delta y_t], \tag{26}$$

where $\mathbf{M}_z$ is the projection matrix $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. But we have just seen that $e_{bt}(\theta_{xy})'\mathbf{Z} = 0 \ \forall \theta_{xy}$, hence equation (26) simplifies to

$$\widehat{\beta_{yx0}} = [e_{bt}(\theta_{xy})'\Delta x_t]^{-1}[e_{bt}(\theta_{xy})'\Delta y_t]. \tag{27}$$

In other words, the two-stage least squares estimator for β_{yx0}, and hence the long-run multiplier θ_{yx}, depends on two key elements: the correlations of the estimated residual from the previous equation, equation (19), with both Δx_t and Δy_t. This is because each equation in a SVAR possesses many regressors in common. Since the "extra" instrument $e_{bt}(\theta_{xy})$ in the second equation is the residual from the first equation, it is by construction orthogonal to the other instruments in the second equation. It then follows that the two-stage least squares estimator for β_{yx0} depends only on the correlations of this residual with Δx_t and Δy_t as shown by (27). To see that certain identification schemes may be problematic, define θ^*_{xy} such that $e_{bt}(\theta^*_{xy})'\Delta x_t = 0$. Then, as long as $e_{bt}(\theta_{xy})'\Delta y_t$ remains finite, $\widehat{\beta_{yx0}}$ diverges when $\theta_{xy} \to \theta^*_{xy}$. In more standard IV settings, this result would not emerge. Residuals from other equations would not generally be used as regressors, and hence parameter estimates would depend on more than one correlation.

To determine the exact value of the problematic identifying restriction, θ^*_{xy}, given the data under consideration, it suffices to take the transpose of equation (21), post-multiply the result by Δx_t, and set it to zero to yield

$$\theta^*_{xy} = \frac{\Delta x_t' \mathbf{W} \Delta x_t}{\Delta y_t' \mathbf{W} \Delta y_t}, \text{ where } \mathbf{W} = \mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}\mathbf{X}' - \text{I}. \tag{28}$$

To continue with our discussion, observe from equations (22) and (27) that

$$\widehat{\beta_{yx0}} - \beta_{yx0} = [e_{bt}(\theta_{xy})'\Delta x_t]^{-1}[e_{bt}(\theta_{xy})'\epsilon_{at}]. \tag{29}$$

Therefore a lower bound for the variance of the two-stage least squares estimator $\widehat{\beta_{yx0}}$ is given by

$$\text{var }(\widehat{\beta_{yx0}}) = \sigma^2_{\epsilon_a}[e_{bt}(\theta_{xy})'\Delta x_t]^{-1}[e_{bt}(\theta_{xy})'e_{bt}(\theta_{xy})][e_{bt}(\theta_{xy})'\Delta x_t]^{-1\prime}, \tag{30}$$

where $\sigma^2_{\epsilon_a} = E(\epsilon^2_{at})$.[4] As $\theta_{xy} \rightarrow \theta^*_{xy}$, this variance diverges at the squared rate of that at which $\widehat{\beta_{yx0}}$ itself diverges. Taken together, equations (27) and (30) tell us that for identification strategies in a neighborhood of θ^*_{xy}, it is not possible to obtain a meaningful estimate of β_{xy0}. Both its estimator as well as associated confidence interval become arbitrarily large.

The above analysis has been numerical in nature in order to make clear the source of identification failure in SVAR estimation. One may wonder further, however, about the relationship between the distributional properties of $\widehat{\beta_{yx0}}$ and the identification restriction θ_{xy}. The questions of statistical inference and asymptotic distribution can be answered to some degree, it turns out, as a special case of the analysis carried out by Staiger and Stock (1993). Their analysis indicates that conventional asymptotic inference procedures are no longer valid when $e_{bt}(\theta_{xy})$ is weakly related to Δx_t in a regression of Δx_t on its instruments.[5]

Since residuals are recursively used as instruments in the estimation of SVARs, the "validity" of the estimation procedure implicitly depends on the nature of the identifying restrictions adopted. That is, the strength of the instruments is contingent upon the identification scheme. Some structural economic models may then be impossible to investigate empirically within the confines of a just-identified SVAR. In particular, as long as an identification strategy generates a small correlation between a recursively estimated residual and the variable it is meant to instrument for in the subsequent equation, coefficient estimates will lose their standard distributional properties.

[4] This is only a lower bound since $e_{bt}(\theta_{xy})$ is a generated regressor and therefore possesses some variation not accounted for in equation (30).

[5] See Appendix.

An Illustrative Example

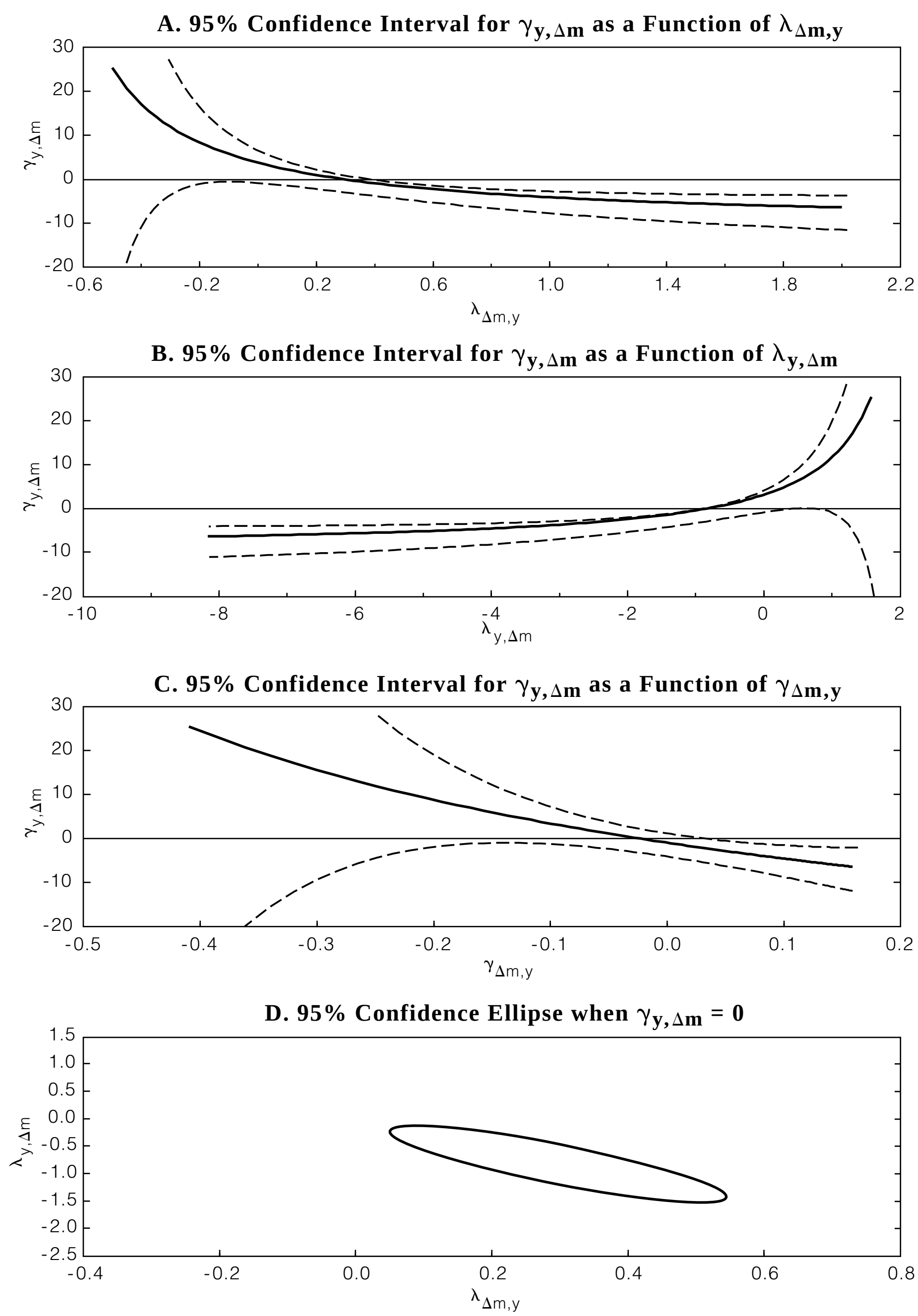
Although the analysis in this section has been carried out with a long-run identifying restriction in mind, the arguments above are also relevant in settings incorporating short-run identifying restrictions. As an example, consider a recent paper on long-run neutrality by King and Watson (1997). The authors estimate a bivariate system in output and money in order to test long-run money neutrality. In doing so, they recognize the importance of considering alternative identifying restrictions for robustness. A subset of their results are reproduced in Figure 1. In panel A of Figure 1, King and Watson (1997) report point estimates and confidence intervals for the hypothesis of long-run superneutrality when the short-run elasticity of money demand with respect to output is allowed to vary. Observe that as this value approaches -0.2, both the coefficient estimate for long-run superneutrality and its confidence intervals begin to blow up. In a similar fashion, panel C shows long-run superneutrality results under various assumptions with respect to the long-run response of money to exogenous permanent shifts in the level of output. Here, $\gamma_{\Delta m,y}$ corresponds to θ_{xy} so that in our notation, Δx_t is the money variable, while y_t is the output variable. As in the case where a short-run identifying restriction was considered, the estimate for long-run superneutrality and its associated confidence intervals start to diverge as $\gamma_{\Delta m,y}$ approaches -0.35. Thus, it should be clear that in looking for robustness across different identification schemes, one may be confronted with cases where the SVAR methodology cannot be meaningfully implemented.

At this stage, there remains at least one other obvious issue of interest. In our context, there may exist a plausible range of identifying restrictions in θ_{xy} for which the residual $e_{bt}(\theta_{xy})$ is, in fact, a proper instrument. If this were the case, one would naturally wonder whether comparative dynamic response estimates are sensitive to the particular identifying restriction imposed upon the system. The next section provides an example of interpretation ambiguities associated with precisely this issue.

4. INTERPRETING STRUCTURAL VARS: TECHNOLOGY SHOCKS AND AGGREGATE EMPLOYMENT FLUCTUATIONS

One topic of considerable interest in macroeconomics is the relationship between technology shocks and aggregate fluctuations in employment. Real business cycle models typically predict that technological innovations raise the level of employment. This result reflects the increase in the marginal productivity of labor associated with the positive technology shock when labor supply is relatively less variable. In a recent paper, however, Gali (1996) suggests that this feature of real business cycle models does not hold empirically. By using a bivariate SVAR in labor productivity and employment, he is able to show

Figure 1 Money Growth and Output



that technology shocks appear to induce a persistent *decline* in employment. Furthermore, labor productivity *increases* temporarily in response to demand shocks.

To motivate the identification of the particular SVAR he uses, Gali (1996) suggests a stylized model whose key features are monopolistic competition, predetermined prices, and variable effort.[6] In such a framework, a positive technology shock enhances labor productivity while leaving aggregate demand unchanged due to sticky prices. Employment must therefore fall. In addition, a positive demand shock would be met by a higher level of "unobserved" effort as well as higher "measured" employment. Given a strong enough effort response, labor productivity would temporarily rise. Formally, the structure of Gali's (1996) model implies that employment evolves according to

$$\Delta h_t = \Theta_{h\eta}(L)\eta_t + \Theta_{h\xi}(L)\xi_t + (1-L)\Phi_{h\eta}(L)\eta_t + (1-L)\Phi_{h\xi}(L)\xi_t, \tag{31}$$

where η_t and ξ_t denote money growth and technology shocks respectively. Here, money growth shocks are associated with the management of aggregate demand by the monetary authority and hence serve as a proxy for demand shocks. Since technology shocks induce a persistent decline in employment, we have $\Theta_{h\xi}(1) < 0$. Similarly, labor productivity is given by

$$\Delta q_t = \Theta_{q\eta}(L)\eta_t + \Theta_{q\xi}(L)\xi_t + (1-L)\Phi_{q\eta}(L)\eta_t + (1-L)\Phi_{q\xi}(L)\xi_t, \tag{32}$$

with $\Theta_{q\eta}(0) + \Phi_{q\eta}(0) > 0$ to capture the contemporaneous positive effect of a demand shock on labor productivity. As in Section 2, this system of equations can be summarized as

$$T(L)Y_t = \epsilon_t, \tag{33}$$

where $Y_t = (\Delta h_t, \Delta q_t)'$, $\epsilon_t = (\eta_t, \xi_t)'$, and

$$T(L) = \begin{bmatrix} \Theta_{h\eta}(L) + (1-L)\Phi_{h\eta}(L) & \Theta_{h\xi}(L) + (1-L)\Phi_{h\xi}(L) \\ \Theta_{q\eta}(L) + (1-L)\Phi_{q\eta}(L) & \Theta_{q\xi}(L) + (1-L)\Phi_{q\xi}(L) \end{bmatrix}^{-1}. \tag{34}$$

The key identifying restriction that Gali (1996) imposes upon the dynamics of his system is that demand shocks do not have a permanent effect on labor productivity. In terms of our earlier notation, we have

$$T(1) = \begin{bmatrix} 1-\theta_{hh} & -\theta_{hq} \\ -\theta_{qh} & 1-\theta_{qq} \end{bmatrix} = \begin{bmatrix} \Theta_{h\eta}(1) & \Theta_{h\xi}(1) \\ \Theta_{q\eta}(1) & \Theta_{q\xi}(1) \end{bmatrix}^{-1}, \tag{35}$$

with

$$\Theta_{q\eta}(1) = \theta_{qh} = 0. \tag{36}$$

[6] For the details of the model, refer to Gali (1996).

Figure 2 plots impulse response functions for the bivariate SVAR we have just described. The data comprise the log of hours worked in the nonfarm business sector as well as gross domestic product (in 1987 dollars), less gross domestic product in the farm sector. The log of productivity was hence computed as the log of gross domestic product, less the log of hours worked. Four lags were used in estimation and the sample period covers 1949:1 to 1992:4. As in Gali (1996), observe that the structural response of employment to a positive technology shock is negative, both in the short and long run. Furthermore, this is true even within a 90 percent confidence interval.[7] Note also that the contemporaneous response of productivity to a demand shock is positive and, by construction, eventually vanishes. Of course, since we have used data that is very similar to that used in the original study, these results are hardly surprising. However, Gali (1996) argues that since these estimates seem to hold for the majority of G7 countries, the impact "of technology shocks yields a picture which is hard to reconcile with the prediction of (real business cycle) models." This statement makes it clear that, among other results, the persistent employment decline in response to a technology shock is implicitly interpreted as a stylized fact. As we know, however, Gali's (1996) estimates derive from his choice of identification scheme; deviations from that scheme must be considered in order to decide what constitutes a stylized fact.

Alternative Identification Strategies

There are several different ways to think about Gali's (1996) initial SVAR set-up. First, supposing that aggregate demand shocks account for more than just money growth shocks, demand shocks may have a permanent impact on productivity. For instance, a permanent increase in taxes in a real business cycle model would yield an increase in the steady-state ratio of employment to capital. Given a standard production function with constant returns to scale, this increase in the ratio of labor to capital would necessarily be accompanied by a fall in labor productivity. This would invalidate the restriction that $\Theta_{q\eta}(1) = \theta_{qh} = 0$. Moreover, since θ_{qh} represents the long-run elasticity of productivity with respect to employment, it might not be unreasonable to expect that $\theta_{qh} < 0$. Figure 3 shows the impulse response functions that result in Gali's (1996) framework when θ_{qh} is set to -0.5. Under this alternative identification strategy, the response of employment to a technology shock is no longer negative. In fact, both the short- and long-run responses of employment are now positive. By comparing Figures 2b and 3b, observe that this latter result seems to hold even when standard errors are taken into account. That is,

[7] To construct the standard error bands, Monte Carlo simulations were done using draws from the normal distribution for each of the two structural innovations. One thousand Monte Carlo draws were carried out in each case.

Figure 2 Identification Assumption: Demand Shocks Have No Long-Run Impact on Productivity

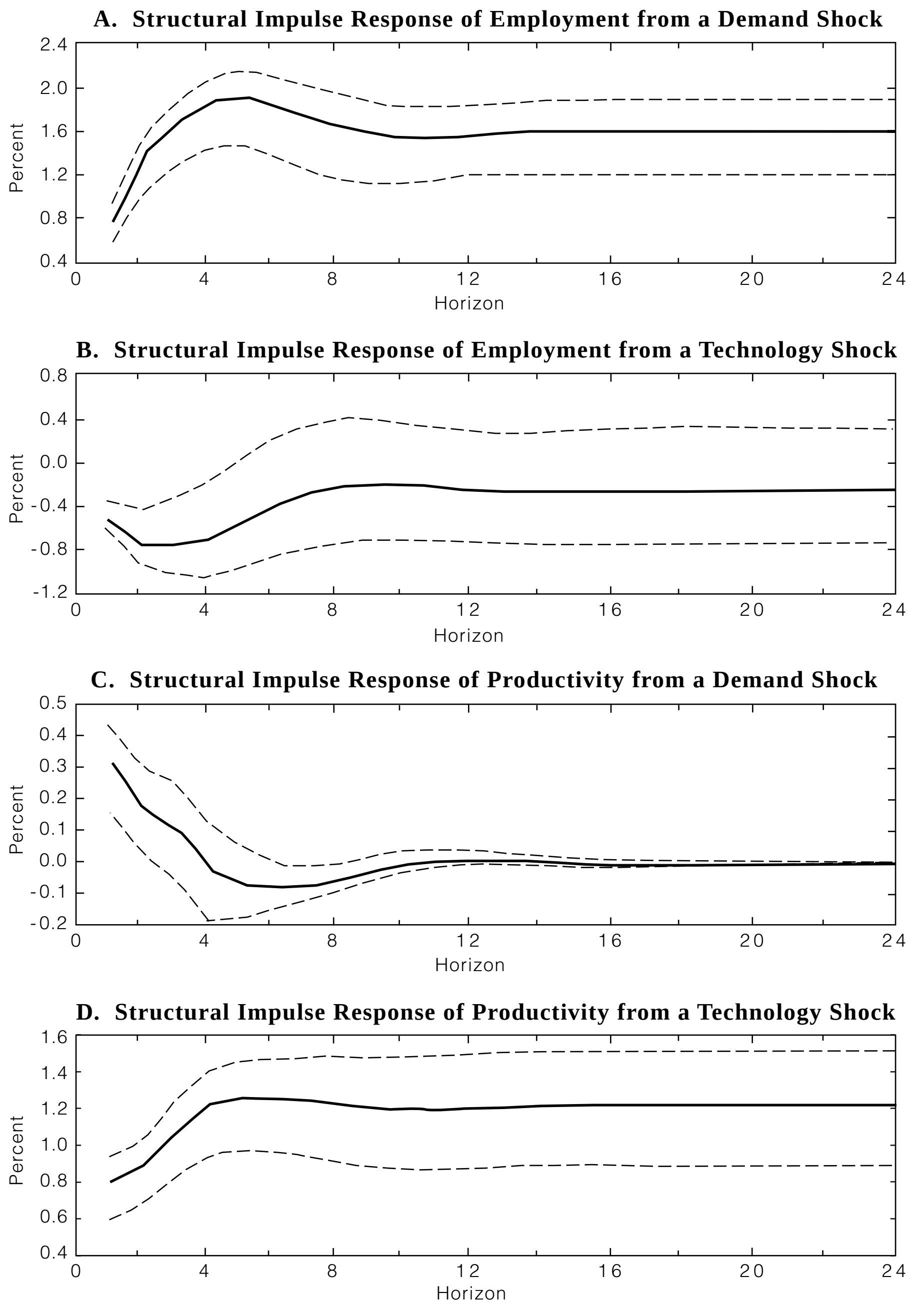
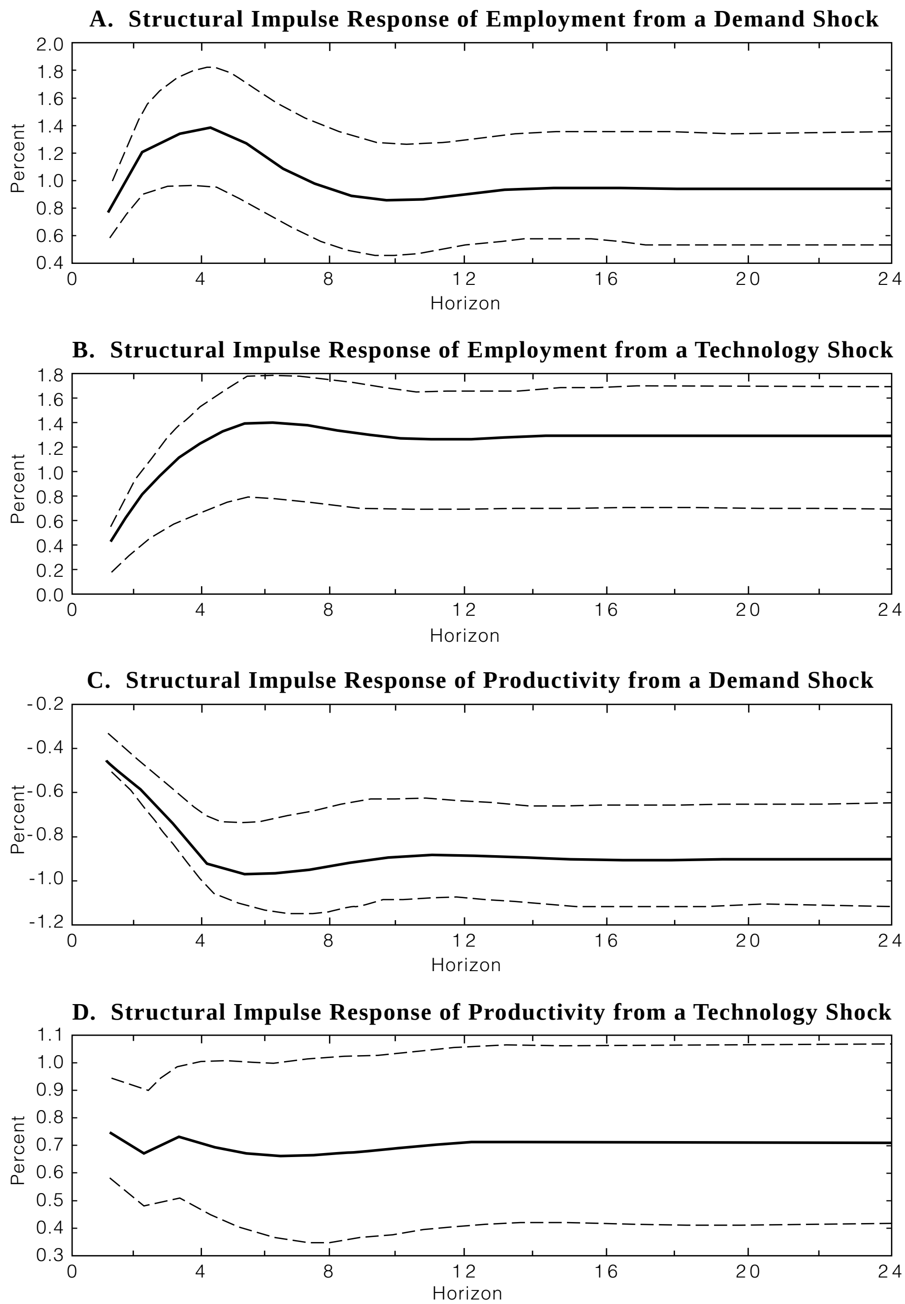


Figure 3 Identification Assumption: Demand Shocks Have a Negative Long-Run Impact on Productivity



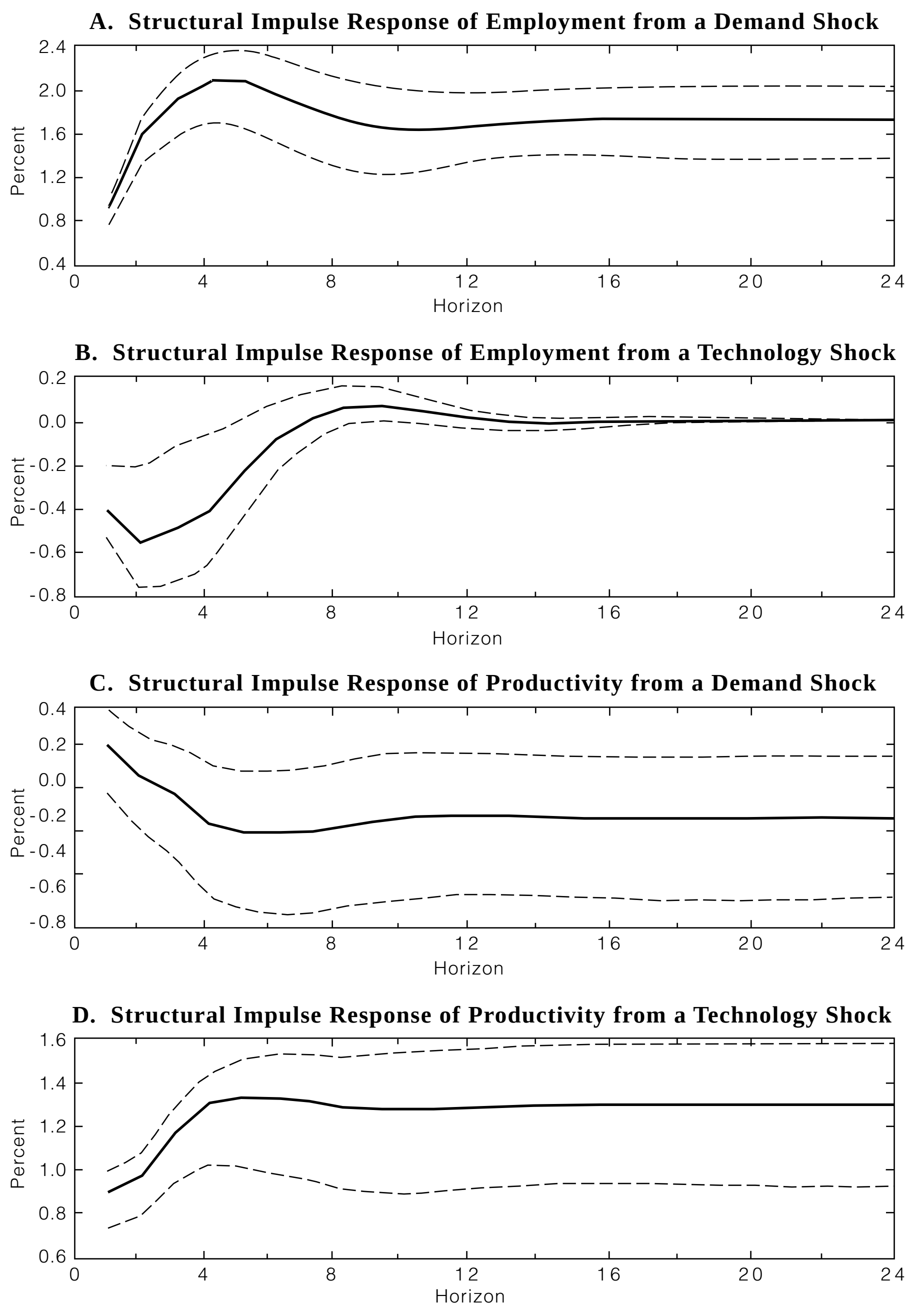
there is little overlap of the corresponding confidence intervals. Moreover, the contemporaneous effect of a demand shock on productivity is no longer positive but negative as shown in panel C. Viewed in this light, the dynamic response estimates initially reported in Gali (1996) may appear somewhat fragile. In particular, his contention that the data does not coincide with the predictions of real business cycle models does not necessarily hold.

In Figure 4, we show the results obtained when Gali's (1996) SVAR is identified using yet a third alternative. In this case, we require that technology shocks not have a long-run impact on employment. In terms of equation (35), this implies that $\Theta_{h\xi}(1) = \theta_{hq} = 0$. This identifying restriction is used by Shapiro and Watson (1988). It also emerges as a steady-state result in a real business cycle model when utility is logarithmic in consumption and leisure. Under this parameterization for utility, the income and substitution effects resulting from a positive technology shock cancel out, leaving labor supply unchanged in the steady state. (See King, Plosser, and Rebelo [1988].) Note in panel C of Figure 4 that under this third alternative, the long-run response of productivity to a demand shock is negative, which provides further evidence against Gali's (1996) initial identifying restriction. As already noted, this result is also consistent with a permanent increase in taxes in a real business cycle framework. Put another way, when one identifies Gali's (1996) bivariate system in a way that is consistent with the steady state generated by a standard real business cycle model, the empirical findings generated by the SVAR are consistent with the predictions of that model.

Of course, that is not to say that real business cycle models represent a more compelling framework when gauged against the data. The empirical results reported by Gali (1996) are themselves consistent with the theoretical model he uses to identify his SVAR. It is simply that in this case, what can be read from the data can vary sharply with one's prior beliefs concerning the theoretical nature of the data-generating mechanism.

While we have just shown that some of the key results in Gali (1996) are sensitive to the way one thinks about the long-run impact of various demand or supply shocks, this is not always the case. Observe that the structural impulse response of employment to a demand shock is similar in both direction and magnitude across Figures 2, 3, and 4. This is also true for the structural impulse response of productivity to a technology shock. Since these latter results emerge across estimated systems, that is, across systems with varying identifying restrictions, they may be reasonably considered stylized facts.

Figure 4 Identification Assumption: Technology Shocks Have No Long-Run Impact on Employment



5. SUMMARY AND CONCLUSIONS

We have investigated the extent to which identification issues can matter when using SVARs to characterize data. Although the main focus was on the estimation of bivariate systems, it should be clear that most of the above analysis applies to larger systems as well.

At a purely mechanical level, the source of the problem lies with the recursive use of an estimated residual as an instrument. The assumption made in SVAR estimation that the structural disturbances be uncorrelated is not sufficient to guarantee a proper estimation procedure. One must also pay attention to the degree of correlation between the estimated residual and the endogenous variable it is meant to be instrumenting for. This observation has long been made for simultaneous equations systems; and in this sense, it is important not to lose sight of the fact that SVARs are in effect a set of simultaneous equations.

At another level, we have also seen that even when the residual from a previously estimated equation is a valid instrument, SVARs can yield ambiguous results. This is the case even when confidence intervals are taken into account as in the bivariate example in hours and productivity. In that case, it was unclear whether employment responded positively or negatively, both in the short and long run, in response to a technology shock. Therefore, there may be a sense in which SVARs can fail in a way that is reminiscent of the Cooley and Leroy (1985) critique. In reduced-form VARs, different results emerge when alternative methods of orthogonalization of the error terms are adopted. In structural VARs, the results can now be directly contingent upon specific identifying restrictions. In effect, these are two facets of the same problem.

We have also seen in our example that certain results may be relatively robust with respect to the particular identification strategy of interest. For example, the response of productivity to a technology shock was estimated to be positive in both the short and long run across varying systems. Thus, two conclusions ultimately emerge from this investigation. First, special emphasis should be given to the derivation of identifying restrictions. The proper use of SVARs is contingent upon such restrictions and the case of identification failure cannot be ruled out a priori. Second, sensitivity analysis can be quite helpful in gaining a sense of the range of dynamics consistent with a given set of data. Assessing such a range seems an essential step in establishing stylized facts.

APPENDIX

This appendix derives the asymptotic distribution of $\widehat{\beta_{yx0}}$ in the text. This derivation is based on Staiger and Stock (1993). In the estimation of equation (22), suppose that the relationship that ties Δx_t to its instruments can be described as

$$\Delta x_t = \mathbf{Z}\alpha + e_{bt}(\theta_{xy})\alpha_{xe} + \nu_t, \tag{A1}$$

where ν_t is uncorrelated with ϵ_{at}. Furthermore, let us consider the set of identifying restrictions $\Pi_{\theta_{xy}}$ for which $\alpha_{xe} = N^{-1/2}g(\theta_{xy})$, where N is the sample size of our dataset and $g(\theta_{xy})$: $\Pi_{\theta_{xy}} \rightarrow \Re$. In other words, $\Pi_{\theta_{xy}}$ denotes a set of identifying restrictions for which the instrument $e_{bt}(\theta_{xy})$ is only weakly related to the endogenous variable Δx_t in the local to zero sense; the coefficient α_{xe} goes to zero as the sample size itself becomes arbitrarily large. To proceed with the argument, rewrite equation (29) as

$$\begin{aligned}
&\widehat{\beta_{yx0}} - \beta_{yx0} = \\
&\quad [(N^{-1/2}\Delta x_t' e_{bt}(\theta_{xy}))(N^{-1}e_{bt}(\theta_{xy})' e_{bt}(\theta_{xy}))(N^{-1/2}e_{bt}(\theta_{xy})'\Delta x_t)]^{-1} \\
&\quad [(N^{-1/2}\Delta x_t' e_{bt}(\theta_{xy}))(N^{-1}e_{bt}(\theta_{xy})' e_{bt}(\theta_{xy}))(N^{-1/2}e_{bt}(\theta_{xy})'\epsilon_{at})].
\end{aligned} \tag{A2}$$

Given the assumptions embodied in (A1), it follows that

$$\begin{aligned}
N^{-1/2}\Delta x_t' e_{bt}(\theta_{xy}) &= N^{-1/2}[\alpha' Z' + \alpha_{xe} e_{bt}(\theta_{xy}) + \nu_t']e_{bt}(\theta_{xy}) \\
&= N^{-1}e_{bt}(\theta_{xy})' e_{bt}(\theta_{xy})g(\theta_{xy}) + N^{-1/2}\nu_t' e_{bt}(\theta_{xy}).
\end{aligned} \tag{A3}$$

Under suitable conditions, the first term in the above equation will converge to some constant almost surely as the sample size becomes large. The second term, on the other hand, will converge asymptotically to a normal distribution by the Central Limit Theorem. Therefore, although the coefficient on the relevant instrument, $e_{bt}(\theta_{xy})$, in the first-stage equation converges to zero, if the rate of convergence is slow enough, the right-hand side of equation (A2) will not diverge asymptotically. Nevertheless, in this case, the two-stage least squares estimator $\widehat{\beta_{yx0}}$ is asymptotically distributed as a ratio of quadratic forms in two jointly distributed normal variables. Hence, for identification strategies that belong to the set $\Pi_{\theta_{xy}}$, conventional asymptotic inference procedures will fail. In fact, in the so-called leading case where $g(\theta_{xy}) = 0$, Phillips (1989), Hillier (1985), and Staiger and Stock (1993) point out that $\widehat{\beta_{yx0}}$ asymptotically possesses a t distribution.

We now provide a sketch of the basic arguments. To this end, we assume that the following moment conditions are satisfied. The notation "$\rightarrow^p$" and "$\Rightarrow$" denote convergence in probability and convergence in distribution respectively.

(a) $(N^{-1}\mathrm{X}'\mathrm{X},\ N^{-1}\mathrm{Z}'\mathrm{X},\ N^{-1}\mathrm{X}'\Delta x_t,\ N^{-1}\mathrm{Z}'\Delta x_t,\ N^{-1}\mathrm{X}'\Delta y_t,\ N^{-1}\mathrm{Z}'\Delta y_t) \rightarrow^p$

$$(\Sigma_{XX},\ \Sigma_{ZX},\ \Sigma_{X\Delta x_t},\ \Sigma_{Z\Delta x_t},\ \Sigma_{X\Delta y_t},\ \Sigma_{Z\Delta y_t})$$

(b) $(N^{-1}\Delta x_t'\Delta x_t,\ N^{-1}\Delta y_t'\Delta y_t,\ N^{-1}\Delta x_t'\Delta y_t) \rightarrow^p (\Sigma_{\Delta x_t\Delta x_t},\ \Sigma_{\Delta y_t\Delta y_t},\ \Sigma_{\Delta x_t\Delta y_t})$

(c) $(N^{-1/2}\nu_t'\Delta x_t,\ N^{-1/2}\nu_t'\Delta y_t,\ N^{-1/2}\nu_t'\mathrm{X},\ N^{-1/2}\epsilon_{at}'\Delta x_t,\ N^{-1/2}\epsilon_{at}'\Delta y_t,$
$N^{-1/2}\epsilon_{at}'\mathrm{X}) \Rightarrow (\Psi_{\nu_t\Delta x_t},\ \Psi_{\nu_t\Delta y_t},\ \Psi_{\nu_t X},\ \Psi_{\epsilon_a\Delta x_t},\ \Psi_{\epsilon_a\Delta y_t},\ \Psi_{\epsilon_a X})$.

Note two particular points embodied in assumptions (a) through (c). First, assumptions (a) and (b) would naturally hold under standard conditions governing stationarity and ergodicity of the variables in the reduced form. Second, since these are primary assumptions, they do not depend on the identifying restriction θ_{xy}. It now remains to specify the asymptotic properties of three terms in (A2) and (A3), namely $N^{-1}e_{bt}(\theta_{xy})'e_{bt}(\theta_{xy})$, $N^{-1/2}\nu_t'e_{bt}(\theta_{xy})$, and $N^{-1/2}e_{bt}(\theta_{xy})'\epsilon_{at}$, to determine the asymptotic behavior of $\widehat{\beta_{yx0}(\theta_{xy})} - \beta_{yx0}(\theta_{xy})$ when $\theta_{xy} \in \Pi_{\theta_{xy}}$. Let us then examine each of these terms in turn.

Recall from equation (21) that

$$e_{bt}(\theta_{xy}) = (\Delta x_t - \theta_{xy}\Delta y_t) - \mathrm{X}(\mathrm{Z}'\mathrm{X})^{-1}\mathrm{Z}'(\Delta x_t - \theta_{xy}\Delta y_t).$$

It follows that $N^{-1}e_{bt}(\theta_{xy})'e_{bt}(\theta_{xy})$ is quadratic in θ_{xy}. Therefore, under assumptions (a) and (b), $N^{-1}e_{bt}(\theta_{xy})'e_{bt}(\theta_{xy}) \rightarrow^p \Sigma(\theta_{xy})$ uniformly, where $\Sigma(\theta_{xy})$ also depends on Σ_{XX}, Σ_{ZX}, etc. Next, consider $N^{-1/2}\nu_t'e_{bt}(\theta_{xy})$. We have

$$N^{-1/2}\nu_t'e_{bt}(\theta_{xy}) = N^{-1/2}[\nu_t'\Delta x_t - \theta_{xy}\nu_t'\Delta y_t - \nu_t'\mathrm{X}(\mathrm{Z}'\mathrm{X})^{-1}\mathrm{Z}'(\Delta x_t - \theta_{xy}\Delta y_t)],$$

which is linear in θ_{xy}. Therefore $N^{-1/2}\nu_t'e_{bt}(\theta_{xy}) \Rightarrow \Psi_{\nu_t}(\theta_{xy})$ uniformly, where $\Psi_{\nu_t}(\theta_{xy}) = \Psi_{\nu_t\Delta x_t} - \theta_{xy}\Psi_{\nu_t\Delta y_t} - \Psi_{\nu_t X}[\Sigma_{ZX}^{-1}\Sigma_{X\Delta x_t} - \theta_{xy}\Sigma_{ZX}^{-1}\Sigma_{Z\Delta y_t}]$. Finally, $N^{-1/2}e_{bt}(\theta_{xy})'\epsilon_{at}$ is given by

$$N^{-1/2}[\Delta x_t'\epsilon_{at} - \theta_{xy}\Delta y_t'\epsilon_{at} - (\Delta x_t' - \theta_{xy}\Delta y_t')\mathrm{Z}(\mathrm{X}'\mathrm{Z})^{-1}\mathrm{X}'\epsilon_{at}],$$

which is also linear in θ_{xy}. Hence, $N^{-1/2}e_{bt}(\theta_{xy})'\epsilon_{at} \Rightarrow \Psi_{\epsilon_{a_t}}(\theta_{xy})$ uniformly, where $\Psi_{\epsilon_{a_t}}(\theta_{xy}) = \Psi_{\epsilon_{a_t}\Delta x_t} - \theta_{xy}\Psi_{\epsilon_{a_t}\Delta y_t} - [\Sigma_{X\Delta x_t}'\Sigma_{ZX}^{-1\prime} - \theta_{xy}\Sigma_{X\Delta y_t}'\Sigma_{ZX}^{-1\prime}]\Psi_{\epsilon_{a_t}X}$. With these results in mind, it follows that $\widehat{\beta_{yx0}(\theta_{xy})}$ converges in distribution to

$$\beta_{yx0}(\theta_{xy}) +$$
$$[(g(\theta_{xy})\Sigma(\theta_{xy})^{1/2} + \Psi_{\nu_t}(\theta_{xy})\Sigma(\theta_{xy})^{-1/2})'(g(\theta_{xy})\Sigma(\theta_{xy})^{1/2} + \Psi_{\nu_t}(\theta_{xy})\Sigma(\theta_{xy})^{-1/2})]^{-1}$$
$$[(g(\theta_{xy})\Sigma(\theta_{xy})^{1/2} + \Psi_{\nu_t}(\theta_{xy})\Sigma(\theta_{xy})^{-1/2})'(\Sigma(\theta_{xy})^{-1/2}\Psi_{\epsilon_{a_t}}(\theta_{xy})].$$

This implies that for identification schemes in $\Pi_{\theta_{xy}}$, the two-stage least squares estimator is not only biased, it is asymptotically distributed as a ratio of quadratic forms in the jointly distributed normal random variables $\Psi_{\nu_t}(\theta_{xy})$ and $\Psi_{\epsilon_{a_t}}(\theta_{xy})$.

REFERENCES

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

Blanchard, Olivier J., and Danny Quah. "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, vol. 79 (September 1989), pp. 655–73.

Blanchard, Olivier J., and Mark W. Watson. "Are Business Cycles All Alike?" in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change,* Chicago: University of Chicago Press, 1984.

Cooley, Thomas F., and Stephen F. Leroy. "Atheoretical Macroeconometrics: A Critique," *Journal of Monetary Economics*, vol. 16 (June 1985), pp. 283–308.

Gali, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" Mimeo, New York University, 1996.

__________. "How Well Does the IS-LM Model Fit Postwar U.S. Data?" *Quarterly Journal of Economics*, vol. 107 (May 1992), pp. 709–38.

Hillier, Grant H. "On the Joint and Marginal Densities of Instrumental Variables Estimators in a General Structural Equation," *Econometric Theory*, vol. 1 (April 1985), pp. 53–72.

King, Robert J., Charles I. Plosser, and Sergio T. Rebelo. "Production, Growth, and Business Cycles: II. New Directions," *Journal of Monetary Economics*, vol. 21 (March 1988), pp. 309–42.

King, Robert J., Charles I. Plosser, James H. Stock, and Mark W. Watson. "Stochastic Trends and Economic Fluctuations," *American Economic Review*, vol. 81 (September 1991), pp. 819–40.

King, Robert J., and Mark W. Watson. "Testing Long-Run Neutrality," Federal Reserve Bank of Richmond *Economic Quarterly,* vol. 83 (Summer 1997), pp. 69–101.

Nelson, Charles R., and Richard Startz. "The Distribution of the Instrumental Variables Estimator and its t-Ratio when the Instrument Is a Poor One," *Journal of Business*, vol. 63 (January 1990), pp. 125–40.

Phillips, Peter. C. "Partially Identified Models," *Econometric Theory*, vol. 5 (August 1989), pp. 181–240.

Shapiro, Mathew D., and Mark W. Watson. "Sources of Business Cycle Fluctuations," *NBER* Working Paper 1246, 1988.

Sims, Christopher. A. “Are Forecasting Models Usable for Policy Analysis?” Federal Reserve Bank of Minneapolis *Quarterly Review*, vol. 10 (Winter 1986), pp. 2–16.

__________. “Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered,” *American Economic Review*, vol. 70 (May 1980a), pp. 250–59.

__________. “Macroeconomics and Reality,” *Econometrica*, vol. 48 (January 1980b), pp. 1–47.

Staiger, Douglas, and James H. Stock. “Instrumental Variables Regressions with Weak Instruments.” Mimeo, Kennedy School of Government, Harvard University, 1993.