When Do Long-Run Identifying Restrictions Give Reliable Results?

Jon FAUST
International Finance Division, Federal Reserve Board, Washington, DC 20551 (faustj@frb.gov)

Eric M. LEEPER
Department of Economics, Indiana University, Bloomington, IN 47405 (eleeper@indiana.edu)

Many recent articles have identified behavioral disturbances in vector autoregressions by imposing restrictions on the long-run effects of shocks. This article demonstrates that this approach will be unreliable unless the underlying economy satisfies three types of strong restrictions. Although many aspects of these issues have been raised before, this article draws out and illustrates the implications for inferences under the long-run scheme. Furthermore, it provides strategies for dealing with the problems.

KEY WORDS: Identification; Long-run restriction; Vector autoregression.

Vector autoregressions (VAR's) have become a popular tool since Sims (1980) labeled as "incredible" the identifying assumptions of large structural econometric models. He argued that many empirical questions could be answered with VAR's identified by more tenable assumptions. Much of the VAR work that followed has focused on finding such assumptions. Initially, Sims proposed imposing a recursive structure on contemporaneous interactions among the variables to identify the model. Because the implied lack of simultaneity generally is not tenable, Bernanke (1986), Blanchard and Watson (1986), and Sims (1986) suggested identifying VAR's by imposing economically plausible restrictions on contemporaneous interactions among variables. This approach has proved useful, but theory often does not provide enough uncontroversial contemporaneous restrictions to identify quantities of interest.

More recently, Blanchard and Quah (1989), King, Plosser, Stock, and Watson (1991), and Shapiro and Watson (1988) advocated basing restrictions on long-run neutrality properties. For example, in many models a shock to the level of the money supply has no long-run effect on output, but a shock to output may affect the long-run level of money. Because many economists find such restrictions plausible a priori, long-run restrictions have been widely used to study the sources of business cycles (Bayoumi and Eichengreen 1993; Rogers and Wang 1993), money supply and demand shocks (Lastrapes and Selgin 1995), and the international transmission of shocks (Hutchison and Walsh 1992; Ahmed, Ickes, Wang, and Yoo 1993).

Sims's critique of large structural models is based on his view that the macroeconomy is a high-dimensional system with rich dynamics and complicated feedbacks among the variables. This view also has strong implications for structural VAR inference, as Sims and others have argued. In this article, we draw on such arguments to assess whether imposing long-run restrictions on small VAR models will give rise to reliable structural inferences.

We discuss three reasons why structural inferences under the long-run scheme may not be reliable. First, the long-run effect of shocks is imprecisely estimated in finite samples, and the long-run identification scheme transfers this imprecision to the estimates of other parameters of the model. We show that, unless strong restrictions are applied, conventional inferences regarding impulse responses will be badly biased in all sample sizes. Two additional reasons come from familiar identification problems inherent in models that aggregate across variables and those that aggregate across time. Although these two issues do not apply exclusively to work under the long-run scheme, we focus on the scheme for concreteness. Each of the three issues has been raised before. This article draws out the implications of these theoretical issues, illustrates their importance, and provides strategies for dealing with the problems.

1. IDENTIFYING VAR'S USING LONG-RUN RESTRICTIONS

This section lays out the basic issues of identification in VAR's and describes the long-run identification scheme. If $X_t = (X_{1t}, \ldots, X_{nt})'$ is covariance stationary, then, ignoring deterministic components, it has a Wold representation,

$$X_t = F(L)u_t. \tag{1}$$

The disturbance term, u_t, has zero mean and is serially uncorrelated with covariance matrix $E[u_t u_t'] = \Sigma$ for all t. The term $F(L)$ is an $(n \times n)$ matrix whose typical element, $f_{ij}(L)$, is a polynomial in the lag operator: $f_{ij}(L) = \sum_{k=0}^{\infty} f_{ijk} L^k$, and $L^k X_t = X_{t-k}$. In the Wold representation, $F_0 = I$, the identity matrix.

If $F(L)$ is invertible, there is also a VAR representation, $R(L)X_t = u_t$, where $R(L) = F(L)^{-1}$ and $R_0 = I$. This representation is known as the *reduced form*, and the Wold representation is the *final form*. In this article we only consider structures with invertible moving average representations; thus, none of our results stem from the nonfundamental representations that Lippi and Reichlin (1993) studied.

An observationally equivalent representation of the process (1) can be formed by taking any nonsingular matrix A_0 and writing

$$X_t = F(L)A_0A_0^{-1}u_t = A(L)\varepsilon_t, \tag{2}$$

where $A(L) = F(L)A_0$ and $\varepsilon_t = A_0^{-1}u_t$. The matrix of coefficients on L^0 in $A(L)$ is A_0. Corresponding to each A_0 is a different *structure* consistent with the final form and a different set of structural shocks, ε_t.

Identification requires choosing n^2 elements of A_0. The first n restrictions are normalizations, fixing the units for each equation. Typically the standard deviations of the shocks are normalized to 1. An additional $n(n-1)/2$ restrictions come from the assumption that the structural shocks are mutually uncorrelated. Together, these restrictions on A_0 imply

$$A_0^{-1}\Sigma A_0^{-1'} = I. \tag{3}$$

1.1 The Long-Run Identification Scheme

Blanchard and Quah (1989), King, Plosser, Stock, and Watson (1991), and Shapiro and Watson (1988) suggested that some or all of the restrictions required to complete the identification could come from long-run neutrality properties. For example, suppose that one believes that nominal shocks have no long-run effect on output. If output is the ith variable in a VAR and the jth shock is nominal, then the neutrality restriction can be written $a_{ij}(1) = \sum_{k=0}^{\infty} a_{ijk} = 0$, or

$$[F(1)A_0]_{i,j} = 0. \tag{4}$$

The restrictions discussed so far identify the shocks only up to a sign transformation. The identification is completed using a priori views about the sign of the impacts or the long-run effects of the structural shocks.

1.2 Illustration: The Blanchard–Quah Model

Throughout the article we use the bivariate model of Blanchard and Quah as an illustration. In this model $X_t = (Y_t, U_t)'$, where Y is the growth rate of gross domestic product (GDP) and U is the unemployment rate among males 20 years or older. The data are quarterly from 1948:2 to 1992:4. Following Blanchard and Quah, output growth has had means extracted for the period through 1973:4 and from 1974:1 onward; unemployment has been linearly detrended. The VAR has eight lags. Three identifying restrictions come from normalizing and orthogonalizing the shocks, and the
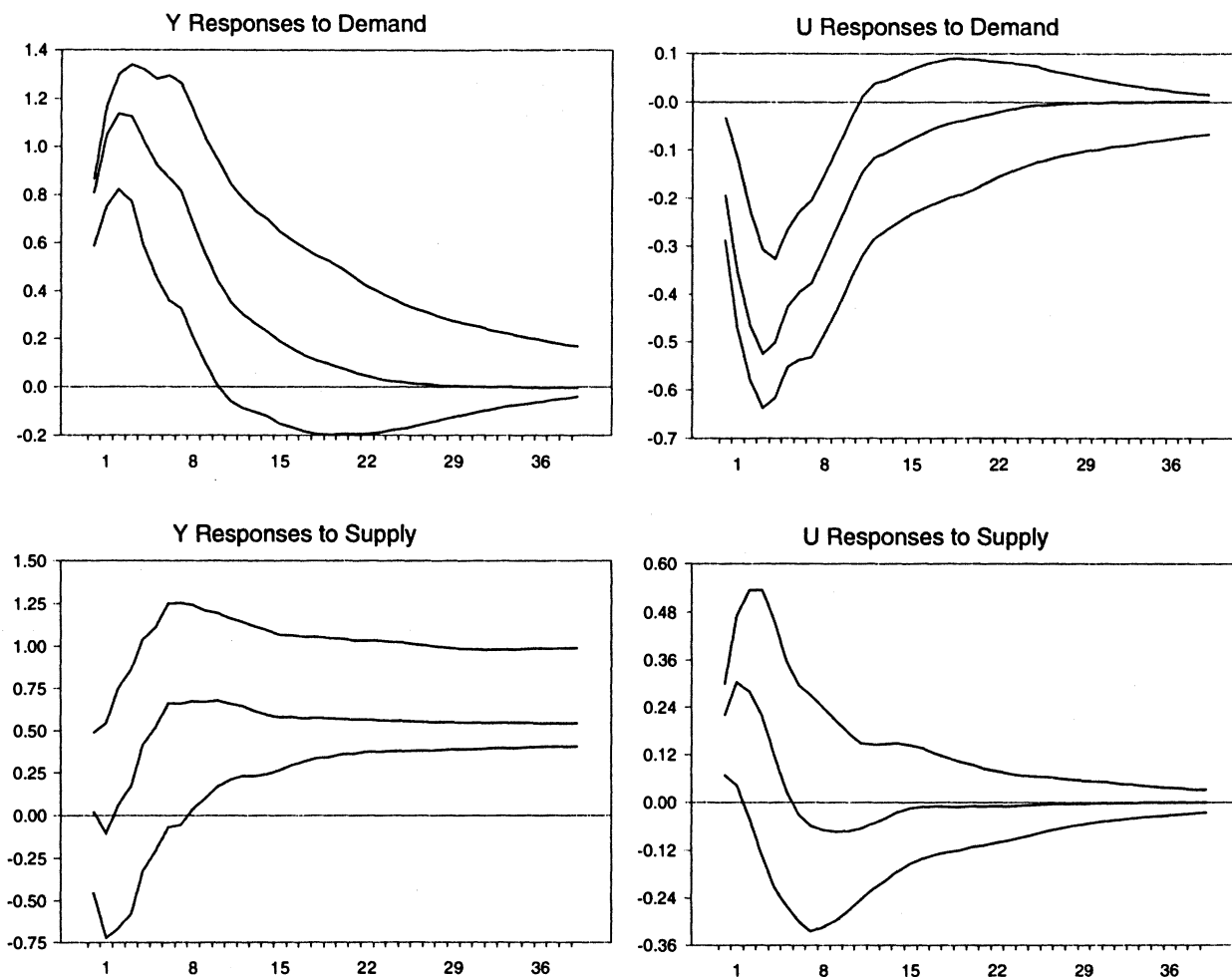

Figure 1. The Vertical Axes Measure the Log of Real GDP or the Rate of Unemployment; the Horizontal Axes Denote Quarters Following the Shock. Point estimates lie inside the empirical 5th and 95th percentile bands taken from 10,000 replications using the Bayesian Monte Carlo procedure in RATS.

Quarter	Y error			U error		
	Lower	Point est.	Upper	Lower	Point est.	Upper
1	52.8	99.9	100.0	2.4	43.9	93.6
2	51.3	99.4	99.6	5.8	53.8	96.3
3	52.2	99.5	99.4	11.1	63.3	97.2
4	52.0	99.0	99.2	16.6	71.0	96.7
8	36.4	84.7	94.9	34.1	82.9	93.4
12	27.1	73.2	91.5	38.7	84.1	92.4
24	17.7	55.3	80.8	40.1	84.3	92.3
40	12.5	42.6	67.6	40.1	84.3	92.3

NOTE: The lower and upper bounds are as in the notes to Figure 1.

final restriction comes from the assumption that nominal shocks have no long-run effect on the level of output. If output is the first variable and the second shock is the nominal disturbance, this implies that $a_{12}(1) = 0$.

Under these assumptions neither shock affects the unemployment rate in the long run, so the assumptions are consistent with a natural rate of unemployment. Positive nominal shocks can shift out aggregate demand and raise output in the short run. The long-run aggregate supply curve is vertical: Output ultimately returns to its original level with either an increase in prices or an inward shift in demand.

Our estimated impulse response functions for supply and demand shocks are nearly identical to those reported by Blanchard and Quah (Fig. 1). Positive nominal shocks have the familiar hump-shaped effect on the level of GDP, peaking after a few quarters and dying out after five years. The output effect of supply shocks grows for two years, then stabilizes at a permanently higher level. The point estimates of the forecast error decompositions (Table 1) imply that demand shocks are the dominant source of output fluctuations for horizons as long as six years. The remainder of the article discusses ways to assess the reliability of structural conclusions such as these and uses this output–unemployment (YU) model to illustrate the issues.

2. LONG–RUN RESTRICTIONS IN FINITE DATA

The long-run restriction, (4), is implemented based on the estimated long-run effects matrix, $\hat{F}(1)$. The reliability of the resulting structural conclusions rests on the quality of the VAR estimate of $F(1)$. This section discusses problems associated with inference under the long-run scheme arising from well-known problems with estimating $F(1)$.

2.1 Problems With Inference Regarding F(1)

VAR work generally relies on weak assumptions on the underlying model generating the data. Often no explicit assumptions are made about the underlying model beyond invertibility of $F(L)$ and summability of the coefficients of $f_{ij}(L)$: $\sum_{k=0}^{\infty} |f_{ijk}| < \infty$.

There is a long literature demonstrating that, if the maintained model imposes only weak dynamic restrictions, then estimates of $F(1)$ are unreliable. Sims (1972) showed that, if the only restriction on $F(L)$ is that the coefficients of each element are summable, then one cannot even form asymptotically correct confidence statements about the value of

$F(1)$ [Blanchard and Quah (1989) noted this problem]. Under certain standard restrictions (e.g., see Hannan 1970), one can form consistent and asymptotically normal estimates of $F(1)$. It is tempting to assume that the asymptotic normal distribution can form the basis for inference about $F(1)$. Following Sims, however, Faust (1996a) showed that, under the standard assumptions giving rise to consistent, asymptotically normal estimates, one still cannot form valid confidence intervals for $F(1)$. This issue is closely related to the literature on the *near observational equivalence* of difference-stationary and trend-stationary processes (Christiano and Eichenbaum 1990; Cochrane 1991; Blough 1992; Faust 1996b).

The source of problems with inferences about $F(1)$ is quite complex, getting one into, as Sims (1972, p. 174) put it, "deep mathematical waters." We state a limited result regarding the long-run scheme that, we believe, communicates the most important aspects of the problem in this context.

Frequently in the VAR literature, we are interested in learning about impulse response functions. It is conventional to provide standard errors on impulse response functions and interpret them as confidence intervals. As Sims and Zha (1995) noted, the classical confidence-interval interpretation is generally not correct when there is uncertainty about unit roots. In this section, we demonstrate a problem with inference about impulse responses under the long-run scheme that occurs even if the assumption of stationarity of all the variables is correct.

Returning to the bivariate, YU model, assume that both the output growth rate and the unemployment rate are stationary. The identified impulse response function is obtained by estimating the final form, $\hat{F}(L)$, and then choosing $\hat{A}_0$ to orthogonalize the errors and to satisfy the long-run restriction: $[\hat{F}(1)\hat{A}_0]_{12} = 0$. This gives the identified moving average polynomial, $\hat{A}(L) = \hat{F}(L)\hat{A}_0$.

Suppose we want to test the hypothesis that the response of the ith variable to the jth shock at lag k is equal to 0: $H_0: a_{ijk} = 0$. The impulse response a_{ijk} is the kth coefficient of $a_{ij}(L)$. Under the long-run scheme, every test of this hypothesis has some very undesirable properties:

Proposition 1. Any test of $H_0: a_{ijk} = 0$ has significance level greater than or equal to maximum power.

Proofs are in the Appendix. The term *maximum power* in this proposition means the largest rejection probability attainable under the test, considering all models for which $a_{ijk} \neq 0$.

What does this result mean? Suppose we limit consideration to tests with a fixed significance level of, say, 5% so that the test rejects H_0 no more than 5% of the time when it is true. The proposition says that the test also rejects the null hypothesis no more than 5% of the time when it is false: The false rejection rate for some models must be greater than the best true rejection rate.

Note further that the rejection probability when H_0 is false must remain less than 5% *no matter what the sample size.* More data does not increase the power of the test. Generally in econometrics, we work with consistent tests—that is, tests that reject a false null with probability 1 in large

samples. Proposition 1 establishes that there are no consistent tests of H_0 [Faust 1996b gave the analogous result for unit-root testing].

The simplest reason for this result is that the estimate of $\hat{F}(1)$ is very uncertain, even in large samples. The long-run restriction transfers this uncertainty to all the coefficients of the impulse response function.

To understand the basics of the proof, take a process $X_t = A(L)\varepsilon_t$, $E\varepsilon_t\varepsilon_t' = I$, that satisfies the long-run scheme and for which $a_{ijk} \neq 0$. Suppose that the test in question has maximum power against this model, rejecting it with greater probability than any other model. Call the rejection probability β. The proof proceeds by altering the process for X_t in two steps. These alterations leave the rejection probability unchanged but result in a process consistent with the null hypothesis and the long-run scheme.

First, alter the X_t process so that it is consistent with H_0. Specifically, form the model $Z_t = B(L)e_t$, where $B(L) = A(L)B$, $e_t = B^{-1}\varepsilon_t$, where B is chosen such that $b_{ijk} = 0$ and $B^{-1}B^{-1'} = I$. Because this model is observationally equivalent to the original, the test must reject H_0 with probability β when this model is true. Although the new model satisfies H_0, it does not satisfy the long-run scheme: $B(1)_{12} = [A(1)B]_{12} \neq 0$. The second alteration makes the process consistent with H_0 and the long-run restriction. Construct $W_t = C(L)e_t$, forming $C(L)$ by taking $B(L)$ and subtracting $B(1)_{12}/m$ from m coefficients of $b_{12}(L)$ (starting with coefficient $k+1$). Now $C(1)_{12} = 0$, so the process for W_t satisfies H_0 and the long-run restriction. The only remaining step is to show that (for sufficiently large m) this last alteration did not alter the rejection probability. The intuition is that, for large m, we have modified m coefficients of $B(L)$ by a tiny amount in forming $C(L)$. The effect of these tiny alterations on the finite-sample behavior of W_t will be negligible, and the test will reject the process for W_t with probability β, just as it does the process for Z_t. Overall, for each process that violates H_0, there is a process consistent with H_0 that the test rejects just as often: size is less than or equal to power.

Given the analogy between tests and confidence intervals, the proposition implies that any valid, say, 95% interval for a_{ijk} must contain 0 at least 95% of the time. It is important to note that conventional methods for computing standard errors on impulse responses—for example, the asymptotic normal approximation and the Bayesian Monte Carlo methods—generally will not reflect the true uncertainty even in large samples. (This applies to the intervals we report in the figures and tables; these are reported for comparability to earlier work.) These methods condition on the VAR specification, including maximum lag length, and thereby eliminate from the outset the vast majority of models consistent with the maintained model. Similarly, most standard tests of $a_{ijk} = 0$ based on asymptotic critical values will not have the proper size.

There is no known way to compute valid critical values for tests of H_0 or meaningful confidence intervals for work under the long-run scheme. The clearest solution is to impose further restrictions.

2.2 Resolving the Problems

There are two obvious sources of additional restrictions. First, one can maintain the long-run restrictions but place sufficient restrictions on the form of $F(L)$ such that estimates of $F(1)$ with meaningful confidence intervals can be obtained. Second, one can give up long-run neutrality restrictions in favor of traditional short-run restrictions.

Sims (1971, 1972) and Faust (1996a,b) discussed the sorts of restrictions that allow meaningful inference regarding $F(1)$. The simplest solution they demonstrate is to assume that the model driving the data is a VAR with known maximum lag order, K. There is surely some K large enough to accommodate most models of interest. In any finite sample, however, as K grows, the confidence intervals for estimates of $F(1)$ grow, as do the confidence intervals for impulse response estimates under the long-run scheme. Thus, this approach will be most useful when one can impose a priori that the true model is exactly a VAR with an order that is small relative to the sample size. Just how small the order of the VAR must be would have to be discovered by simulation.

An alternative approach to strengthening the long-run restriction is to restate the restriction as a finite-horizon restriction, imposing, for example, that the effect of some shock is 0 at 40 quarters and beyond. This will involve overidentifying restrictions. (A different approach is to impose that the effect of the shock is 0 at 40 quarters but to impose no restriction for periods after 40 quarters. This is not consistent with long-run neutrality, however, because this restriction only requires that the net effect of the shock *cross* 0 at 40 quarters, saying nothing about the effect at longer horizons.)

The second suggestion is to identify the model using standard short-run (finite-horizon) restrictions and then to use the long-horizon responses as an informal diagnostic. King and Watson (1992) used this approach to study the relation between money and output. They identified the money supply shock using a broad range of identifying assumptions on impact elasticities and then examined the implied response of output to a nominal shock at various horizons. Having identified the nominal shock using finite-horizon restrictions, one is free to assess whether the speed with which and the extent to which the effect of the nominal shock on output dies out is consistent with one's view of money neutrality.

Overall, complicated inference problems arise when impulse responses are identified under the long-run scheme. These can be avoided by imposing strong a priori restrictions on the lag length of the underlying model or on the horizon at which the effect of the shock goes to 0. An alternative is to use short-run restrictions and use the moderate-to-long horizon properties of the model as an informal diagnostic.

3. THE PROBLEM POSED BY MULTIPLE SHOCKS

The VAR methodology is usually applied in low-dimensional models, so the identified shocks must be viewed as aggregates of a larger number of underlying

shocks. In the YU model, for example, the estimated supply shock must combine oil shocks, labor-supply shocks, and productivity shocks. As Blanchard and Quah (1989) noted this poses a problem. In general, *even if none of the underlying demand shocks affects output in the long run,* the long-run scheme will commingle the underlying demand and supply shocks in both of the estimated disturbances, invalidating the economic interpretation. Blanchard and Quah (1989, p. 670) provided a theorem specifying when this commingling will not occur and concluded on a priori grounds that the scheme gives reasonable and useful results for the YU model.

In this section, we provide a stronger basis for assessing the usefulness of the long-run approach. In particular, we present a simple reformulation of Blanchard and Quah's theorem, draw out its theoretical implications, and recommend an approach for empirically assessing the implications.

3.1 Conditions for Valid Shock Aggregation

We posit a model driven by many shocks and show the conditions under which a low-dimensional VAR identified with the long-run scheme will correctly identify the shocks. Suppose that the true system for $X_t = (Y_t, U_t)'$ is driven by m shocks ($m > 2$):

$$X_t = \tilde{A}(L)\tilde{\varepsilon}_t, \tag{5}$$

where $\tilde{A}(L)$ is $(2 \times m)$, $E[\tilde{\varepsilon}_t] = 0$, and $E[\tilde{\varepsilon}_t\tilde{\varepsilon}_t'] = I$. Assume that each shock is either a supply shock or a demand shock and that $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{1t}', \tilde{\varepsilon}_{2t}')'$, where $\tilde{\varepsilon}_{1t}$ and $\tilde{\varepsilon}_{2t}$ are $(m_1 \times 1)$ and $(m_2 \times 1)$ vectors of supply shocks and demand shocks, respectively, with $m = m_1 + m_2$.

Assume that the long-run restriction holds so that no demand shock in the underlying model has a long-run effect on output: $\tilde{A}_{1j}(1) = 0$ if $j > m_1$. If there were just one shock of each type, this restriction, combined with orthogonality restriction, would identify both shocks.

Now consider the two-shock representation of (5) consistent with the long-run scheme. There must be a two-shock final form, $X_t = F(L)u_t$, where u_t is (2×1). Take this form and choose A_0 to satisfy the long-run scheme, giving

$$X_t = A(L)\varepsilon_t, \tag{6}$$

where $E\varepsilon_t\varepsilon_t' = I$.

Equation (6) is observationally equivalent to the representation of X_t in (5). The question is: When will the identified, two-shock representation in (6) give an aggregate demand shock involving only underlying demand shocks and an aggregate supply shock made up only of underlying supply shocks? In general, the two identified shocks will be a mixture of all the underlying shocks. Part (1) of Proposition 2 is a reformulation of Blanchard and Quah's theorem and states when the two categories of shocks will be properly sorted out, but the timing of shocks will be distorted. Part (2) states when both the shock categories and timing of shocks will be preserved.

Proposition 2. Given the structure (5) and the two-shock representation (6), (1) for $j = 1, 2$, the shock ε_{jt}, will be a

linear function only of the elements of $\tilde{\varepsilon}_{js}$, $s \leq t$, only if

$$\tilde{A}(L) = \Gamma(L) \quad D(L),$$
$$(2 \times 2) \quad (2 \times m)$$

and $D(L)$ is block diagonal when partitioned conformably with the shock categories (i.e., $D_{ij}(L) = 0$ if $i = 1$ and $j > m_1$ or if $i = 2$ and $j \leq m_1$), and (2) for $j = 1, 2$, the shock ε_{jt}, will be a linear function only of $\tilde{\varepsilon}_{jt}$, only if part (1) holds and $D(L) = D$, a block diagonal matrix of scalars.

Although Proposition 2 has important implications, the proof is trivial. Equate the two representations (5) and (6): $\tilde{A}(L)\tilde{\varepsilon}_t = A(L)\varepsilon_t$. If each ε_{jt} is a linear combination of the underlying category j shocks at and before t, then by definition there is a $D(L)$ that is block diagonal such that $\varepsilon_t = D(L)\tilde{\varepsilon}_t$. Thus, $\tilde{A}(L)\tilde{\varepsilon}_t = A(L)D(L)\tilde{\varepsilon}_t$, which is the required result. When $D(L) = D$, part (2) follows.

The part (2) conditions, under which the categories and the timing of shocks will be preserved, require that each underlying shock of a given type affects the economy in the same way up to a scale factor. This is implausible in most cases. In general, when estimating low-dimensional models, the dynamic response of the economy to any particular underlying demand shock will differ from the estimated impulse response for the *aggregate shock*.

For an economic interpretation of the part (1) restriction, consider shutting down all the shocks in the model except the kth supply shock. If the part (1) conditions hold, we can write, $Y_t = \gamma_{11}(L)d_{1k}(L)\tilde{\varepsilon}_{kt}$ and $U_t = \gamma_{21}(L)d_{1k}(L)\tilde{\varepsilon}_{kt}$, implying

$$U_t = \gamma_{21}(L)^{-1}\gamma_{11}(L)Y_t. \tag{7}$$

Because $d_{1k}(L)$ drops out, (7) holds for *every* supply shock: The response of U to every supply shock can be expressed as a single distributed lag on Y. An analogous result holds for the demand shock.

It is easy to check whether the implication in (7) holds in any theoretical model motivating the empirical work. For example, suppose that the lead coefficient of $(\gamma_{21}(L)^{-1}\gamma_{11}(L))$ is negative. This implies that every supply shock that increases output growth on impact must also decrease the unemployment rate. This may be inconsistent with standard reasoning about productivity and labor-supply shocks. A productivity shock probably will increase output and decrease unemployment on impact. An exogenous increase in female labor-force participation, however, may increase employment and output, but if employment initially increases by less than the labor force, unemployment rises with output.

3.2 Empirical Plausibility of the Part (1) Restrictions

The problem in Proposition 2 stems from the fact that the underlying model has more sources of shocks than does the estimated model. One solution, advocated by Sims (1980), would be to estimate larger systems. When one wishes to maintain the convenience of small models, however, an alternative approach is to check for consistency of results across various small models.
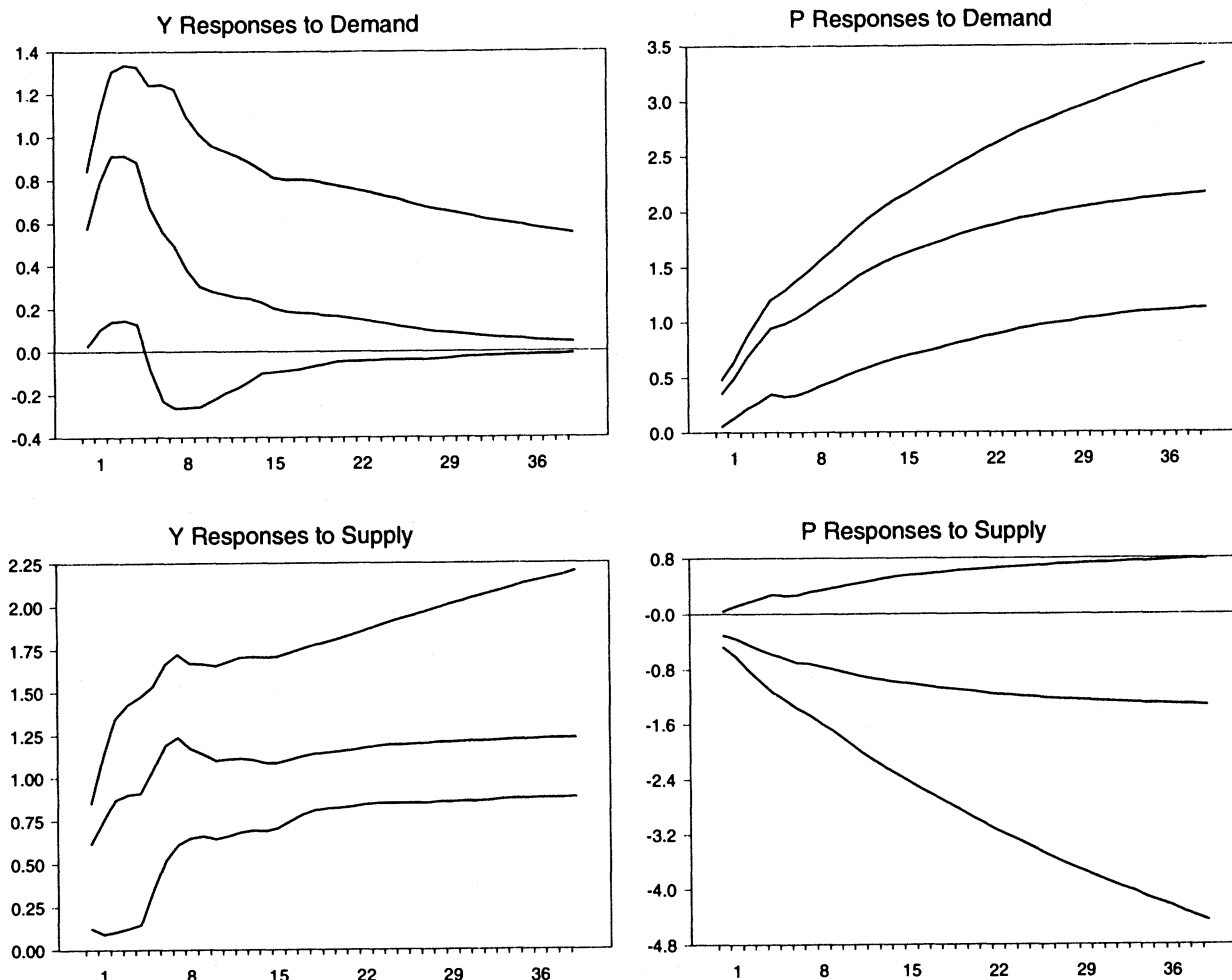
Figure 2. The Vertical Axes Measure the Log of Real GDP or the Log of the GDP Deflator; the Horizontal Axes Denote Quarters Following the Shock. See note to Figure 1.

Often there are several different variables on which a given analysis could be based. Instead of the YU model of supply and demand shocks, an output–price-level model might be equally appealing. The aggregation theorem may hold in neither, one, or both of these systems. If it holds for both, the supply shocks estimated in both models will be uncorrelated asymptotically with the demand shocks in both models, and vice versa. If the supply shock from one model is correlated with the demand shock from the other model, there is clear evidence that one or both of the models have commingled the underlying supply and demand shocks.

We compare the YU results to those from an output and inflation (YP) model similar to that of Bayoumi and Eichengreen (1993). Inflation is measured as the quarterly growth rate of the GDP deflator, with means extracted for the pre-

Table 2. Contemporaneous Correlation Among the Shocks in the YU and YP Models

	Demand shocks		Supply shocks	
Shocks	YU-D	YP-D	YU-S	YP-S
YU-D	1.00	.65	.00	.56
YP-D		1.00	−.13	.00
YU-S			1.00	.20
YP-S				1.00

and post-1974 periods. The VAR is estimated with eight lags. The identifying assumption is that demand shocks do not affect output in the long run. The impulse responses for the YP model appear reasonable, having the correct signs and the familiar humped shape (Fig. 2). There is clear evidence, however, that the YU and YP models have aggregated the underlying demand and supply shocks differently. The YU and YP supply shocks are weakly correlated with each other, and the demand shocks are only moderately correlated (Table 2). More troubling is the fact that the YP supply shock is more highly correlated with the YU demand shock (.56) than with the YU supply shock (.20). The probable commingling of underlying demand and supply shocks has important interpretational implications: Demand shocks in the YP model have a much smaller role in determining output over horizons of two to five years than in the YU model (Table 3).

The models' estimated shocks shed some light on the differences in the models. For example, in the first quarter of 1951, a small drop in GDP coincided with a big spike in inflation, which was influenced by the Korean War. The YP model finds an extraordinary—three-standard-deviation—outward shift in demand and a similar magnitude inward shift in supply. The YU model does not have to account for inflation fluctuations and reports nothing extraordinary.

Table 3. Percent of Forecast Error Variance Due to Demand in the YP Model

	Y error			P error		
Quarter	Lower	Point est.	Upper	Lower	Point est.	Upper
1	1.9	46.4	96.8	3.7	56.5	98.5
2	2.7	50.2	97.5	5.8	62.2	98.6
3	3.0	51.0	97.5	7.1	66.3	98.8
4	3.0	50.9	97.5	8.0	68.3	98.9
8	3.3	37.2	90.4	8.3	69.4	99.0
12	2.8	27.6	80.5	8.6	70.2	99.0
24	1.6	15.7	57.4	9.0	71.6	99.2
40	1.0	9.4	36.6	9.2	72.2	99.3

NOTE: The lower and upper bounds are as in the notes to Figure 1.

This example illustrates a simple consistency check on the results from small models. In the example, we see no strong a priori or empirical grounds for selecting between the two models and conclude that neither provides a reliable basis for structural inference.

4. THE PROBLEM POSED BY HIGH–FREQUENCY FEEDBACKS

In VAR modeling, long-run restrictions typically are coupled with the assumption that structural shocks are orthogonal. The standard justification for this assumption is the view that the shocks originate in behaviorally distinct sectors of the economy. Even if this view is correct, however, the assumption of orthogonality may be inappropriate in time-aggregated or infrequently sampled data. For example, suppose we view the stock-market drop in October of 1987 as a supply shock. The Federal Reserve Board reacted within hours, injecting reserves into the banking system. In quarterly data, these two changes will be contemporaneous; sorting out whether this is a nominal or a real shock will be impossible.

There is a long literature on identifying continuous-time models from discrete data (Telser 1967; Sims 1971; Phillips 1973; Geweke 1978; Hansen and Sargent 1991a). Hansen and Sargent (1991b) and Marcet (1991) took up this issue in the context of VAR inference. The general lesson from this work as it applies, say, to the YU model is that, even when the continuous-time structural shocks are uncorrelated, both shocks identified in discrete data under the orthogonality assumption will commingle the underlying supply and demand shocks. The results of this section are essentially a special case of Marcet's and Hansen and Sargent's results. This case comes from assuming that there is some discrete frequency of observation high enough that the identification assumptions hold.

4.1 Time Aggregation and Identification

Following the logic of Section 3, we specify a model operating at high frequency and check when a model estimated on time-aggregated data will get the right answer. Although the following results would hold for any higher and lower frequency of observation, we consider quarterly and annual frequencies. Thus, take a quarterly version of the YU model that is driven by two shocks and satisfies the

long-run restriction

$$\tilde{X}_t = \tilde{A}(L)\tilde{\varepsilon}_t, \qquad t = 1, 2, \ldots, 4T, \qquad (8)$$

with $[\tilde{A}(1)]_{12} = 0$ and $E[\tilde{\varepsilon}_t\tilde{\varepsilon}_t'] = I$.

Now assume that the data, X_t, are annual observations. In particular, X_t is a linear function of the four values of $\tilde{X}_t$ making up the year, $X_t = M(L)\tilde{X}_t$, $t = 4, 8, \ldots, 4T$, where $M(L)$ is diagonal, of order 3, and known. The model for the observed data is $X_t = M(L)\tilde{A}(L)\tilde{\varepsilon}_t$, $t = 4, 8, \ldots, 4T$. This expresses X_t in terms of all the underlying quarterly shocks. There must be a final form for X_t with only one shock per year that can be written $X_t = F(L^4)u_t$, $t = 4, 8, \ldots, 4T$. Application of the long-run scheme leads to

$$X_t = A(L^4)\varepsilon_t, \qquad t = 4, 8, \ldots, 4T, \qquad (9)$$

with $E[\varepsilon_t\varepsilon_t'] = I$ and $a_{21}(1) = 0$.

As in Section 3, the central question regards when the annual demand shock will be a linear function only of the underlying quarterly demand shocks, and similarly for supply. The relevant condition is derived just as in Proposition 2.

Proposition 3. Given the quarterly structure (8) and annual representation (9), (1) for $j = 1, 2$, the annual shock ε_{jt} will be a linear function only of the quarterly shock $\tilde{\varepsilon}_{js}$, $s = t, t - 1, \ldots$, only if $M(L)\tilde{A}(L) = \Gamma(L^4)D(L)$, where $D(L)$ is diagonal, and (2) for $j = 1, 2$, the shock ε_{jt} will be a linear function only of $\tilde{\varepsilon}_{js}$, $s = t, t - 1, \ldots, t - 3$, only if part (1) holds and the diagonal elements of $D(L)$ are of order less than 4.

Once again, part (1) shows when the demand and supply shocks will be properly sorted out but the timing will be distorted; part (2) shows when the shock types and shock timing will be preserved.

The simplest case in which the conditions of Proposition 3 are met is when $\tilde{A}(L)$ is diagonal, implying that neither variable Granger-causes the other at the quarterly frequency. Proposition 3 does allow feedbacks, but only of a very limited variety. Following the same procedure used to derive (7), we can show that the response of U_t to any supply shock must satisfy $U_t = \gamma_{21}(L^4)^{-1}\gamma_{11}(L^4)Y_t$, $t = 1, 2, \ldots, 4T$. Although few plausible models would deliver this prediction that the response of quarterly U_t to a supply shock is expressible simply in terms $Y_t, Y_{t-4}, \ldots$, this restriction may approximately hold in practice.

4.2 Assessing the Empirical Relevance of Proposition 3

To assess the practical importance of Proposition 3, we estimate the YU and YP models discussed previously using annual average data. Many of the broad features of the quarterly models carry over to the annual models. In the YU system, much of the forecast-error variance in output is attributed to demand at business-cycle frequencies, but the YP model gives much less importance to demand.

One way to characterize how similarly the quarterly and annual models separate supply and demand is to assume that the estimated quarterly models are correct and to ask how the annual models aggregate the quarterly ones. It follows from the quarterly and annual representations (8) and (9),

that the annual shocks are related to the quarterly shocks by $\varepsilon_t = A(L^p)^{-1}M(L)\tilde{A}(L)\tilde{\varepsilon}_t = Z(L)\tilde{\varepsilon}_t$, $t = 4, 8, \ldots, 4T$. The shock types are commingled if $Z(L)$ is not diagonal, and we can evaluate how close the matrix is to diagonal by substituting the estimated quarterly and annual lag polynomials for $\tilde{A}(L)$ and $A(L)$, respectively.

We summarize the results by reporting the proportion of the overall variance of the annual demand shock and supply shock that is caused by the underlying quarterly demand and supply shocks. In the YU model, the annual supply shock involved substantial commingling, with 27% of the variance of the annual supply shock accounted for by the quarterly demand shock. The annual YU demand shock is somewhat less confounded, with only 9% of the variance caused by the quarterly supply shock. The annual YP model involves almost no commingling of the quarterly shocks: About 95% of the variance of each annual shock is accounted for by the corresponding quarterly shock. This result emerges in part because there are few feedbacks from Y to P (Y does not Granger-cause P in the quarterly data).

Given the ubiquitous feedbacks present in most general equilibrium models and the tight restrictions imposed by Proposition 3, it might seem likely that time aggregation would greatly muddle the results. In the YP and YU models, however, the muddling was only moderate.

5. SUMMARY

The explicit assumptions of the long-run identifying scheme have been viewed as weak and innocuous and, thereby, as protected from the "incredible" label applied to other approaches. This article shows that structural inference under the long-run scheme will be reliable, however, only if the underlying structure being approximated by the VAR satisfies strong dynamic restrictions. The results of this article do not suggest that the long-run scheme should be abandoned. The results do not even provide a clear ranking of the scheme against other VAR, real-business-cycle, or Cowles Commission approaches. The results are further evidence that identification in macroeconomics is a dirty business and that care must be taken to assess the robustness of inference. We provide several approaches to evaluating and improving the robustness of inferences under the long-run scheme.

ACKNOWLEDGMENTS

APPENDIX: PROOFS OF PROPOSITIONS

Proof of Proposition 1. The following sketch fills out two details of the proof in the text. For technical details,

see Faust (1996a,b). First, show that there is a B satisfying

$$B^{-1}B^{-1'} = I \tag{A.1}$$

such that $b_{ijk} = 0$. Define A_k as the matrix of coefficients of L^k in $A(L)$. Now $b_{ijk} = [A_k B]_{ij}$ so that the required restrictions are (A.1) and $[A_k B]_{ij} = 0$. These restrictions form four equations in four unknown elements of B, which can be solved directly.

Now prove that, for any fixed sample size T, the rejection frequency under the process for W_t for sufficiently large m is β. To be clear, define $W_t^m = C^m(L)e_t$, where $C^m(L)$ is the same as $B(L)$ except that $c_{ij}^m(L)$ is modified as described in the text and the index m, giving the number of modified coefficients, is made explicit. Clearly, the process for each W_t^m satisfies all assumptions of the long-run scheme. In forming $C^m(L)$, we alter m coefficients of $B(L)$ by the amount $B(1)_{12}/m$. The sum of the squared alterations in the coefficients is $B(1)_{12}^2/m$, which goes to 0 with m. This implies, for example, by a variant of Bernstein's lemma that the random variable made up of a sample of size T from the W^m process converges in joint distribution to the analogous sample size T random variable from the Z_t process. Under certain standard restrictions, this further implies that the rejection probability of the test when the W^m process is true must converge with m to β, the rejection probability under the Z process.

Proof of Proposition 2. The proof is in the text.

Proof of Proposition 3. Part (1): From the two representations of X_t and assuming the conclusion, we have $M(L)\tilde{A}(L)\tilde{\varepsilon}_t = A(L^4)\varepsilon_t = A(L^4)Z(L)\tilde{\varepsilon}_t$ for some diagonal $Z(L)$. Thus, $M(L)\tilde{A}(L) = A(L^4)Z(L)$, proving the point.

Part (2): Repeat the proof of part (1) with the order of $D(L)$ limited.

REFERENCES

Ahmed, S., Ickes, B. W., Wang, P., and Yoo, B. S. (1993), "International Business Cycles," *American Economic Review*, 83, 335–359.

Bayoumi, T., and Eichengreen, B. (1993), "Shocking Aspects of European Monetary Unification," in *Adjustment and Growth in the European Monetary Union*, eds. F. Torres and F. Giavazzi, Cambridge, U.K.: Cambridge University Press, pp. 193–229.

Bernanke, B. (1986), "Alternative Explorations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–99.

Blanchard, O. J., and Quah, D. (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655–673.

Blanchard, O. J., and Watson, M. W. (1986), "Are Business Cycles All Alike?" in *The American Business Cycle: Continuity and Change*, ed. R. J. Gordon, Chicago: University of Chicago Press, pp. 123–156.

Blough, S. R. (1992), "The Relationship Between Power and Level for Generic Unit Root Tests in Finite Samples," *Journal of Applied Econometrics*, 7, 295–308.

Christiano, L. J., and Eichenbaum, M. (1990), "Unit Roots in Real GNP: Do We Know, and Do We Care?" *Carnegie-Rochester Conference Series on Public Policy*, 32, 7–62.

Cochrane, J. H. (1991), "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control*, 15, 275–284.

Faust, J. (1996a), "Theoretical Confidence Level Problems With Conventional Confidence Intervals for the Spectrum of a Time Series," International Finance Discussion Paper 575, Federal Reserve Board.

—— (1996b), "Near Observational Equivalence and Theoretical Size Problems With Unit Root Tests," *Econometric Theory*, 12, 724–731.

Geweke, J. B. (1978), "Temporal Aggregation in the Multivariate Regression Model," *Econometrica*, 46, 643–662.

Hannan, E. J. (1970), *Multiple Time Series*, New York: Wiley.

Hansen, L. P., and Sargent, T. J. (1991a), "Two Difficulties in Interpreting Vector Autoregressions," in *Rational Expectations Econometrics*, eds. L. P. Hansen and T. J. Sargent, Boulder, CO: Westview Press, pp. 77–120.

—— (1991b), "Identification of Continuous Time Rational Expectations Models From Discrete Time Data," in *Rational Expectations Econometrics*, eds. L. P. Hansen and T. J. Sargent, Boulder, CO: Westview Press, pp. 219–238.

Hutchison, M. M., and Walsh, C. E. (1992), "Empirical Evidence on the Insulation Properties of Fixed and Flexible Exchange Rates," *Journal of International Economics*, 32, 241–263.

King, R. G., Plosser, C. I., Stock, J. H., and Watson, M. W. (1991), "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

King, R. G., and Watson, M. W. (1992), "Testing Long-Run Neutrality," unpublished manuscript, Northwestern University, Dept. of Economics.

Lastrapes, W. D., and Selgin, G. (1995), "The Liquidity Effect: Identifying Short-Run Interest Rate Dynamics Using Long-Run Restrictions," *Journal of Macroeconomics*, 17, 387–404.

Lippi, M., and Reichlin, L. (1993), "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment," *American Economic Review*, 83, 644–652.

Marcet, A. (1991), "Temporal Aggregation of Economic Time Series," in *Rational Expectations Econometrics*, eds. L. P. Hansen and T. J. Sargent, Boulder, CO: Westview Press, pp. 237–282.

Phillips, P. C. B. (1973), "The Problem of Identification in Finite Parameter Continuous-Time Models," *Journal of Econometrics*, 1, 351–362.

Rogers, J. H., and Wang, P. (1993), "Sources of Fluctuations in Relative Prices: Evidence From High Inflation Countries," *Review of Economics and Statistics*, 75, 587–605.

Shapiro, M. D., and Watson, M. W. (1988), "Sources of Business Cycle Fluctuations," in *NBER Macroeconomics Annual 1988*, ed. S. Fischer, Cambridge, MA: MIT Press, pp. 111–148.

Sims, C. A. (1971), "Discrete Approximations to Continuous Time Lag Distributions in Econometrics," *Econometrica*, 39, 545–564.

—— (1972), "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *Journal of the American Statistical Association*, 67, 169–175.

—— (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

—— (1986), "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter, 2–16.

Sims, C. A., and Zha, T. (1995), "Error Bands for Impulse Responses," Working Paper 95–6, Federal Reserve Bank of Atlanta.

Telser, L. G. (1967), "Discrete Samples and Moving Sums in Stationary Stochastic Processes," *Journal of the American Statistical Association*, 62, 484–499.