This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: NBER Macroeconomics Annual 2006, Volume 21

Volume Author/Editor: Daron Acemoglu, Kenneth Rogoff and Michael Woodford, editors

Volume Publisher: The MIT Press

Volume ISBN: 0-262-01239-1; 978-0-262-01239-3

Volume URL: http://www.nber.org/books/acem06-1

Conference Dates: April 7-8, 2006

Publication Date: May 2007

Chapter Title: Assessing Structural VARs

Chapter Author: Lawrence J. Christiano, Martin Eichenbaum, Robert Vigfusson

Chapter URL: http://www.nber.org/chapters/c11177

Chapter pages in book: (1 - 106)

1 Assessing Structural VARs

Lawrence J. Christiano, *Northwestern University, the Federal Reserve Bank of Chicago, and NBER*
Martin Eichenbaum, *Northwestern University, the Federal Reserve Bank of Chicago, and NBER*
Robert Vigfusson, *Federal Reserve Board of Governors*

1 Introduction

Sims's seminal paper *Macroeconomics and Reality* (1980) argued that procedures based on vector autoregression (VAR) would be useful to macroeconomists interested in constructing and evaluating economic models. Given a minimal set of identifying assumptions, structural VARs allow one to estimate the dynamic effects of economic shocks. The estimated impulse response functions provide a natural way to choose the parameters of a structural model and to assess the empirical plausibility of alternative models.[1]

To be useful in practice, VAR-based procedures must have good sampling properties. In particular, they should accurately characterize the amount of information in the data about the effects of a shock to the economy. Also, they should accurately uncover the information that is there.

These considerations lead us to investigate two key issues. First, do VAR-based confidence intervals accurately reflect the actual degree of sampling uncertainty associated with impulse response functions? Second, what is the size of bias relative to confidence intervals, and how do coverage rates of confidence intervals compare with their nominal size?

We address these questions using data generated from a series of estimated dynamic, stochastic general equilibrium (DSGE) models. We consider real business cycle (RBC) models and the model in Altig, Christiano, Eichenbaum, and Linde (2005) (hereafter, ACEL) that embodies real and nominal frictions. We organize most of our analysis around a particular question that has attracted a great deal of attention in the literature: How do hours worked respond to an identified shock? In the case of the RBC model, we consider a neutral shock to technology. In the ACEL model, we consider two types of technology shocks as well as a monetary policy shock.

We focus our analysis on an unavoidable specification error that occurs when the data generating process is a DSGE model and the econometrician uses a VAR. In this case the true VAR is infinite ordered, but the econometrician must use a VAR with a finite number of lags.

We find that as long as the variance in hours worked due to a given shock is above the remarkably low number of 1 percent, VAR-based methods for recovering the response of hours to that shock have good sampling properties. Technology shocks account for a much larger fraction of the variance of hours worked in the ACEL model than in any of our estimated RBC models. Not surprisingly, inference about the effects of a technology shock on hours worked is much sharper when the ACEL model is the data generating mechanism.

Taken as a whole, our results support the view that structural VARs are a useful guide to constructing and evaluating DSGE models. Of course, as with any econometric procedure it is possible to find examples in which VAR-based procedures do not do well. Indeed, we present such an example based on an RBC model in which technology shocks account for less than 1 percent of the variance in hours worked. In this example, VAR-based methods work poorly in the sense that bias exceeds sampling uncertainty. Although instructive, the example is based on a model that fits the data poorly and so is unlikely to be of practical importance.

Having good sampling properties does not mean that structural VARs always deliver small confidence intervals. Of course, it would be a Pyrrhic victory for structural VARs if the best one could say about them is that sampling uncertainty is always large and the econometrician will always know it. Fortunately, this is not the case. We describe examples in which structural VARs are useful for discriminating between competing economic models.

Researchers use two types of identifying restrictions in structural VARs. Blanchard and Quah (1989), Galí (1999), and others exploit the implications that many models have for the long-run effects of shocks.[2] Other authors exploit short-run restrictions.[3] It is useful to distinguish between these two types of identifying restrictions to summarize our results.

We find that structural VARs perform remarkably well when identification is based on short-run restrictions. For all the specifications that we consider, the sampling properties of impulse response estimators are good and sampling uncertainty is small. This good performance obtains even when technology shocks account for as little as 0.5 percent of the variance in hours. Our results are comforting for the vast literature that has exploited short-run identification schemes to identify the dynamic effects of shocks to the economy. Of course, one can question the particular short-run identifying assumptions used in any given analysis. However, our results strongly support the view that if the relevant short-run assumptions are satisfied in the data generating process, then standard structural VAR procedures reliably uncover and identify the dynamic effects of shocks to the economy.

The main distinction between our short and long-run results is that the sampling uncertainty associated with estimated impulse response functions is substantially larger in the long-run case. In addition, we find some evidence of bias when the fraction of the variance in hours worked that is accounted for by technology shocks is very small. However, this bias is not large relative to sampling uncertainty as long as technology shocks account for at least 1 percent of the variance of hours worked. Still, the reason for this bias is interesting. We document that, when substantial bias exists, it stems from the fact that with long-run restrictions one requires an estimate of the sum of the VAR coefficients. The specification error involved in using a finite-lag VAR is the reason that in some of our examples, the sum of VAR coefficients is difficult to estimate accurately. This difficulty also explains why sampling uncertainty with long-run restrictions tends to be large.

The preceding observations led us to develop an alternative to the standard VAR-based estimator of impulse response functions. The only place the sum of the VAR coefficients appears in the standard strategy is in the computation of the zero-frequency spectral density of the data. Our alternative estimator avoids using the sum of the VAR coefficients by working with a nonparametric estimator of this spectral density. We find that in cases when the standard VAR procedure entails some bias, our adjustment virtually eliminates the bias.

Our results are related to a literature that questions the ability of long-run identified VARs to reliably estimate the dynamic response of macroeconomic variables to structural shocks. Perhaps the first critique of this sort was provided by Sims (1972). Although his paper was written before the advent of VARs, it articulates why estimates of the sum of regression coefficients may be distorted when there is specification error. Faust and Leeper (1997) and Pagan and Robertson (1998) make an important related critique of identification strategies based on long-run restrictions. More recently Erceg, Guerrieri, and Gust (2005) and Chari, Kehoe, and McGrattan (2005b) (henceforth, CKM) also examine the reliability of VAR-based inference using long-run identifying restrictions.[4] Our conclusions regarding the value of identified VARs differ sharply from those recently reached by CKM. One parameterization of the RBC model that we consider is identical to the one considered by CKM. This parameterization is included for pedagogical purposes only, as it is overwhelmingly rejected by the data.

The remainder of the paper is organized as follows. Section 2 presents the versions of the RBC models that we use in our analysis. Section 3 discusses our results for standard VAR-based estimators of impulse response functions. Section 4 analyzes the differences between short and long-run restrictions. Section 5 discusses the relation between our work and the recent critique of VARs offered by CKM. Section 6 summarizes the ACEL model and reports its implications for VARs. Section 7 contains concluding comments.

2 A Simple RBC Model

In this section, we display the RBC model that serves as one of the data generating processes in our analysis. In this model the only shock that affects labor productivity in the long-run is a shock to technology. This property lies at the core of the identification strategy used by King et al. (1991), Galí (1999) and other researchers to identify the effects of a shock to technology. We also consider a variant of the model which rationalizes short run restrictions as a strategy for identifying a technology shock. In this variant, agents choose hours worked before the technology shock is realized. We describe the conventional VAR-based strategies for estimating the dynamic effect on hours worked of a shock to technology. Finally, we discuss parameterizations of the RBC model that we use in our experiments.

2.1 *The Model*

The representative agent maximizes expected utility over per capita consumption, c_t, and per capita hours worked, l_t:

$$E_0 \sum_{t=0}^{\infty} (\beta(1+\gamma))^t \left[\log c_t + \psi \frac{(1-l_t)^{1-\sigma} - 1}{1-\sigma} \right],$$

subject to the budget constraint:

$$c_t + (1 + \tau_{x,t})\, i_t \leq (1 - \tau_{l,t}) w_t l_t + r_t k_t + T_t,$$

where

$$i_t = (1 + \gamma)\, k_{t+1} - (1 - \delta) k_t.$$

Here, k_t denotes the per capita capital stock at the beginning of period t, w_t is the wage rate, r_t is the rental rate on capital, $\tau_{x,t}$ is an investment tax, $\tau_{l,t}$ is the tax rate on labor income, $\delta \in (0, 1)$ is the depreciation rate on capital, γ is the growth rate of the population, T_t represents lump-sum taxes and $\sigma > 0$ is a curvature parameter.

The representative competitive firm's production function is:

$$y_t = k_t^{\alpha} (Z_t l_t)^{1-\alpha},$$

where Z_t is the time t state of technology and $\alpha \in (0, 1)$. The stochastic processes for the shocks are:

$$\log z_t = \mu_z + \sigma_z \varepsilon_t^z \tag{1}$$

$$\tau_{l,t+1} = (1 - \rho_l)\tau_l + \rho_l \tau_{l,t} + \sigma_l \varepsilon_{t+1}^l$$

$$\tau_{x,t+1} = (1 - \rho_x)\tau_x + \rho_x \tau_{x,t} + \sigma_x \varepsilon_{t+1}^x,$$

where $z_t = Z_t / Z_{t-1}$. In addition, ε_t^z, ε_t^l, and ε_t^x are independently and identically distributed (i.i.d.) random variables with mean zero and unit standard deviation. The parameters, σ_z, σ_l, and σ_x are non-negative scalars. The constant, μ_z, is the mean growth rate of technology, τ_l is the mean labor tax rate, and τ_x is the mean tax on capital. We restrict the autoregressive coefficients, ρ_l and ρ_x, to be less than unity in absolute value.

Finally, the resource constraint is:

$$c_t + (1 + \gamma)\, k_{t+1} - (1 - \delta) k_t \leq y_t.$$

We consider two versions of the model, differentiated according to timing assumptions. In the *standard* or *nonrecursive version*, all time t decisions are taken after the realization of the time t shocks. This is the conventional assumption in the RBC literature. In the *recursive version* of the model the timing assumptions are as follows. First, $\tau_{l,t}$ is observed, and then labor decisions are made. Second, the other shocks are realized and agents make their investment and consumption decisions.

2.2 *Relation of the RBC Model to VARs*

We now discuss the relation between the RBC model and a VAR. Specifically, we establish conditions under which the reduced form of the RBC model is a VAR with disturbances that are linear combinations of the economic shocks. Our exposition is a simplified version of the discussion in Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) (see especially their section III). We include this discussion because it frames many of the issues that we address. Our discussion applies to both the standard and the recursive versions of the model.

We begin by showing how to put the reduced form of the RBC model into a state-space, observer form. Throughout, we analyze the log-linear approximations to model solutions. Suppose the variables of interest in the RBC model are denoted by X_t. Let s_t denote the vector of exogenous economic shocks and let $\hat{k}_t$ denote the percent deviation from steady state of the capital stock, after scaling by Z_t.[5] The approximate solution for X_t is given by:

$$X_t = a_0 + a_1\hat{k}_t + a_2\hat{k}_{t-1} + b_0 s_t + b_1 s_{t-1}, \tag{2}$$

where

$$\hat{k}_{t+1} = A\hat{k}_t + Bs_t. \tag{3}$$

Also, s_t has the law of motion:

$$s_t = Ps_{t-1} + Q\varepsilon_t, \tag{4}$$

where ε_t is a vector of i.i.d. fundamental economic disturbances. The parameters of (2) and (3) are functions of the structural parameters of the model.

The "state" of the system is composed of the variables on the right side of (2):

$$\xi_t = \begin{pmatrix} \hat{k}_t \\ \hat{k}_{t-1} \\ s_t \\ s_{t-1} \end{pmatrix}.$$

The law of motion of the state is:

$$\xi_t = F\xi_{t-1} + D\varepsilon_t, \tag{5}$$

where F and D are constructed from A, B, Q, P. The econometrician observes the vector of variables, Y_t. We assume Y_t is equal to X_t plus iid measurement error, v_t, which has diagonal variance-covariance, R. Then:

$$Y_t = H\xi_t + v_t. \tag{6}$$

Here, H is defined so that $X_t = H\xi_t$, that is, relation (2) is satisfied. In (6) we abstract from the constant term. Hamilton (1994, section 13.4) shows how the system formed by (5) and (6) can be used to construct the exact Gaussian density function for a series of observations, $Y_1, \ldots, Y_T$. We use this approach when we estimate versions of the RBC model.

We now use (5) and (6) to establish conditions under which the reduced form representation for X_t implied by the RBC model is a VAR with disturbances that are linear combinations of the economic shocks. In this discussion, we set $v_t = 0$, so that $X_t = Y_t$. In addition, we assume that the number of elements in ε_t coincides with the number of elements in Y_t.

We begin by substituting (5) into (6) to obtain:

$$Y_t = HF\xi_{t-1} + C\varepsilon_t, \qquad C \equiv HD.$$

Our assumption on the dimensions of Y_t and ε_t implies that the matrix C is square. In addition, we assume C is invertible. Then:

$$\varepsilon_t = C^{-1}Y_t - C^{-1}HF\xi_{t-1}. \tag{7}$$

Substituting (7) into (5), we obtain:

$$\xi_t = M\xi_{t-1} + DC^{-1}Y_t,$$

where

$$M = [I - DC^{-1}H]F. \tag{8}$$

As long as the eigenvalues of M are less than unity in absolute value,

$$\xi_t = DC^{-1}Y_t + MDC^{-1}Y_{t-1} + M^2DC^{-1}Y_{t-2} + \ldots\,. \tag{9}$$

Using (9) to substitute out for ξ_{t-1} in (7), we obtain:

$$\varepsilon_t = C^{-1}Y_t - C^{-1}HF[DC^{-1}Y_{t-1} + MDC^{-1}Y_{t-2} + M^2DC^{-1}Y_{t-3} + \ldots],$$

or, after rearranging:

$$Y_t = B_1Y_{t-1} + B_2Y_{t-2} + \ldots + u_t, \tag{10}$$

where

$$u_t = C\varepsilon_t \tag{11}$$

$$B_j = HFM^{j-1}DC^{-1}, \quad j = 1, 2, \ldots \tag{12}$$

Expression (10) is an infinite-order VAR, because u_t is orthogonal to Y_{t-j}, $j \geq 1$.

Proposition 2.1. *(Fernandez-Villaverde, Rubio-Ramirez, and Sargent) If C is invertible and the eigenvalues of M are less than unity in absolute value, then the RBC model implies:*

- Y_t has the infinite-order VAR representation in (10)
- The linear one-step-ahead forecast error Y_t given past Y_t's is u_t, which is related to the economic disturbances by (11)
- The variance-covariance of u_t is CC'
- The sum of the VAR lag matrices is given by:

$$B(1) \equiv \sum_{j=1}^{\infty} B_j = HF[I - M]^{-1}DC^{-1}.$$

We will use the last of these results below.

Relation (10) indicates why researchers interested in constructing DSGE models find it useful to analyze VARs. At the same time, this relationship clarifies some of the potential pitfalls in the use of VARs. First, in practice the econometrician must work with finite lags. Second, the assumption that C is square and invertible may not be satisfied. Whether C satisfies these conditions depends on how Y_t is defined. Third, significant measurement errors may exist. Fourth, the matrix, M, may not have eigenvalues inside the unit circle. In this case, the economic shocks are not recoverable from the VAR disturbances.[6] Implicitly, the econometrician who works with VARs assumes that these pitfalls are not quantitatively important.

2.3 *VARs in Practice and the RBC Model*

We are interested in the use of VARs as a way to estimate the response of X_t to economic shocks, i.e., elements of ε_t. In practice, macroeconomists use a version of (10) with finite lags, say q. A researcher can estimate $B_1, \ldots, B_q$ and $V = Eu_tu_t'$. To obtain the impulse response functions, however, the researcher needs the B_i's and the column of C corresponding to the shock in ε_t that is of interest. However, to compute the required column of C requires additional identifying assumptions. In practice, two types of assumptions are used. Short-run assumptions take the form of direct restrictions on the matrix C. Long-run assumptions place indirect restrictions on C that stem from restrictions on the long-run response of X_t to a shock in an element of ε_t. In this section we use our RBC model to discuss these two types of assumptions and how they are imposed on VARs in practice.

2.3.1 The Standard Version of the Model The log-linearized equilibrium laws of motion for capital and hours in this model can be written as follows:

$$\log \hat{k}_{t+1} = \gamma_0 + \gamma_k \log \hat{k}_t + \gamma_z \log z_t + \gamma_l \tau_{l,t} + \gamma_x \tau_{x,t}, \tag{13}$$

and

$$\log l_t = a_0 + a_k \log \hat{k}_t + a_z \log z_t + a_l \tau_{l,t} + a_x \tau_{x,t}. \tag{14}$$

From (13) and (14), it is clear that all shocks have only a temporary effect on l_t and $\hat{k}_t$.[7] The only shock that has a permanent effect on labor productivity, $a_t \equiv y_t/l_t$, is ε^z_t. The other shocks do not have a permanent effect on a_t. Formally, this *exclusion restriction* is:

$$\lim_{j\to\infty}[E_t a_{t+j} - E_{t-1}a_{t+j}] = f(\varepsilon_t^z \text{ only}). \tag{15}$$

In our linear approximation to the model solution f is a linear function. The model also implies the *sign restriction* that f is an increasing function. In (15), E_t is the expectation operator, conditional on the information set $\Omega_t = (\log \hat{k}_{t-s}, \log z_{t-s}, \tau_{l,t-s}, \tau_{x,t-s}; s \ge 0)$.

In practice, researchers impose the exclusion and sign restrictions on a VAR to compute ε_t^z and identify its dynamic effects on macroeconomic variables. Consider the $N \times 1$ vector, Y_t. The VAR for Y_t is given by:

$$Y_{t+1} = B(L)Y_t + u_{t+1}, \quad Eu_t u_t' = V, \tag{16}$$

$$B(L) \equiv B_1 + B_2 L + \ldots + B_q L^{q-1},$$

$$Y_t = \begin{pmatrix} \Delta \log a_t \\ \log l_t \\ x_t \end{pmatrix}.$$

Here, x_t is an additional vector of variables that may be included in the VAR. Motivated by the type of reasoning discussed in the previous subsection, researchers assume that the fundamental economic shocks are related to u_t as follows:

$$u_t = C\varepsilon_t, \quad E\varepsilon_t \varepsilon_t' = I, \quad CC' = V. \tag{17}$$

Without loss of generality, we assume that the first element in ε_t is ε_t^z. We can easily verify that:

$$\lim_{j \to \infty} [\tilde{E}_t a_{t+j} - \tilde{E}_{t-1} a_{t+j}] = \tau [I - B(1)]^{-1} C\varepsilon_t, \tag{18}$$

where τ is a row vector with all zeros, but with unity in the first location. Here:

$$B(1) \equiv B_1 + \ldots + B_q.$$

Also, $\tilde{E}_t$ is the expectation operator, conditional on $\tilde{\Omega}_t = \{Y_t, \ldots, Y_{t-q+1}\}$. As mentioned above, to compute the dynamic effects of ε_t^z, we require $B_1, \ldots, B_q$ and C_1, the first column of C.

The symmetric matrix, V, and the B_i's can be computed using ordinary least squares regressions. However, the requirement that $CC' = V$ is not sufficient to determine a unique value of C_1. Adding the exclusion and sign restrictions does uniquely determine C_1. Relation (18) implies that these restrictions are:

$$\text{exclusion restriction: } [I - B(1)]^{-1} C = \begin{bmatrix} \text{number} & \underline{0} \\ \text{numbers} & \text{numbers} \end{bmatrix},$$

where $\underline{0}$ is a row vector and

sign restriction: (1,1) element of $[I - B(1)]^{-1} C$ is positive.

There are many matrices, C, that satisfy $CC' = V$ as well as the exclusion and sign restrictions. It is well-known that the first column, C_1, of each of these matrices is the same. We prove this result here, because elements of the proof will be useful to analyze our simulation results. Let

$$D \equiv [I - B(1)]^{-1} C.$$

Let $S_Y(\omega)$ denote the spectral density of Y_t at frequency ω that is implied by the q^{th}-order VAR. Then:

$$DD' = [I - B(1)]^{-1} V [I - B(1)']^{-1} = S_Y(0). \tag{19}$$

The exclusion restriction requires that D have a particular pattern of zeros:

$$D = \begin{bmatrix} \underset{1\times 1}{d_{11}} & \underset{1\times(N-1)}{0} \\ \underset{(N-1)\times 1}{D_{21}} & \underset{(N-1)\times(N-1)}{D_{22}} \end{bmatrix}$$

so that

$$DD' = \begin{bmatrix} d_{11}^2 & d_{11}D_{21}' \\ D_{21}d_{11} & D_{21}D_{21}' + D_{22}D_{22}' \end{bmatrix} = \begin{bmatrix} S_Y^{11}(0) & S_Y^{21}(0)' \\ S_Y^{21}(0) & S_Y^{22}(0) \end{bmatrix},$$

where

$$S_Y(\omega) \equiv \begin{bmatrix} S_Y^{11}(\omega) & S_Y^{21}(\omega)' \\ S_Y^{21}(\omega) & S_Y^{22}(\omega) \end{bmatrix}.$$

The exclusion restriction implies that

$$d_{11}^2 = S_Y^{11}(0), \quad D_{21} = S_Y^{21}(0) / d_{11}. \tag{20}$$

There are two solutions to (20). The sign restriction

$$d_{11} > 0 \tag{21}$$

selects one of the two solutions to (20). So, the first column of D, D_1, is uniquely determined. By our definition of C, we have

$$C_1 = [I - B(1)]D_1. \tag{22}$$

We conclude that C_1 is uniquely determined.

2.3.2 The Recursive Version of the Model In the recursive version of the model, the policy rule for labor involves $\log z_{t-1}$ and $\tau_{x,t-1}$ because these variables help forecast $\log z_t$ and $\tau_{x,t}$:

$$\log l_t = a_0 + a_k \log \hat{k}_t + \tilde{a}_l \tau_{l,t} + \tilde{a}_z' \log z_{t-1} + \tilde{a}_x' \tau_{x,t-1}.$$

Because labor is a state variable at the time the investment decision is made, the equilibrium law of motion for $\hat{k}_{t+1}$ is:

$$\log \hat{k}_{t+1} = \gamma_0 + \gamma_k \log \hat{k}_t + \tilde{\gamma}_z \log z_t + \tilde{\gamma}_l \tau_{l,t} + \tilde{\gamma}_x \tau_{x,t} + \tilde{\gamma}_z' \log z_{t-1} + \tilde{\gamma}_x' \tau_{x,t-1}.$$

As in the standard model, the only shock that affects a_t in the long run is a shock to technology. So, the long-run identification strategy discussed in section 2.3.1 applies to the recursive version of the model. However, an alternative procedure for identifying ε_t^z applies to this version of the model. We refer to this alternative procedure as the "short-run" identification strategy because it involves recovering ε_t^z using only the realized one-step-ahead forecast errors in labor productivity and hours, as well as the second moment properties of those forecast errors.

Let $u^a_{\Omega,t}$ and $u^l_{\Omega,t}$ denote the population one-step-ahead forecast errors in a_t and $\log l_t$, conditional on the information set, Ω_{t-1}. The recursive version of the model implies that

$$u^a_{\Omega,t} = \alpha_1 \varepsilon_t^z + \alpha_2 \varepsilon_t^l, \qquad u^l_{\Omega,t} = \gamma \varepsilon_t^l,$$

where $\alpha_1 > 0$, α_2, and γ are functions of the model parameters. The projection of $u^a_{\Omega,t}$ on $u^l_{\Omega,t}$ is given by

$$u^a_{\Omega,t} = \beta u^l_{\Omega,t} + \alpha_1 \varepsilon_t^z, \quad \text{where } \beta = \frac{cov(u^a_{\Omega,t}, u^l_{\Omega,t})}{var(u^l_{\Omega,t})}. \tag{23}$$

Because we normalize the standard deviation of ε_t^z to unity, α_1 is given by:

$$\alpha_1 = \sqrt{var(u^a_{\Omega,t}) - \beta^2 var(u^l_{\Omega,t})}.$$

In practice, we implement the previous procedure using the one-step-ahead forecast errors generated from a VAR in which the variables in Y_t are ordered as follows:

$$Y_t = \begin{pmatrix} \log l_t \\ \Delta \log a_t \\ x_t \end{pmatrix}.$$

We write the vector of VAR one-step-ahead forecast errors, u_t, as:

$$u_t = \begin{pmatrix} u_t^l \\ u_t^a \\ u_t^x \end{pmatrix}.$$

We identify the technology shock with the second element in ε_t in (17). To compute the dynamic response of the variables in Y_t to the technology shock we need $B_1, \ldots, B_q$ in (16) and the second column, C_2, of the matrix C, in (17). We obtain C_2 in two steps. First, we identify the technology shock using:

$$\varepsilon_t^z = \frac{1}{\hat{\alpha}_1}(u_t^a - \hat{\beta} u_t^l),$$

where

$$\hat{\beta} = \frac{cov(u_t^a, u_t^l)}{var(u_t^l)}, \quad \hat{\alpha}_1 = \sqrt{var(u_t^a) - \hat{\beta}^2 var(u_t^l)}.$$

The required variances and covariances are obtained from the estimate of V in (16). Second, we regress u_t on ε_t^z to obtain:[8]

$$C_2 = \begin{pmatrix} \dfrac{cov(u^l, \varepsilon^z)}{var(\varepsilon^z)} \\ \dfrac{cov(u^a, \varepsilon^z)}{var(\varepsilon^z)} \\ \dfrac{cov(u^x, \varepsilon^z)}{var(\varepsilon^z)} \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\alpha}_1 \\ \dfrac{1}{\hat{\alpha}_1}(cov(u_t^x, u_t^a) - \hat{\beta} cov(u_t^x, u_t^l)) \end{pmatrix}.$$

2.4 *Parameterization of the Model*

We consider different specifications of the RBC model that are distinguished by the parameterization of the laws of motion of the exogenous shocks. In all specifications we assume, as in CKM, that:

$$\beta = 0.98^{1/4}, \quad \theta = 0.33, \quad \delta = 1 - (1 - .06)^{1/4}, \quad \psi = 2.5, \quad \gamma = 1.01^{1/4} - 1 \quad (24)$$
$$\tau_x = 0.3, \quad \tau_l = 0.242, \quad \mu_z = 1.016^{1/4} - 1, \quad \sigma = 1.$$

2.4.1 Our MLE Parameterizations We estimate two versions of our model. In the *two-shock maximum likelihood estimation (MLE) specification* we assume that $\sigma_x = 0$, so that there are two shocks, $\tau_{l,t}$ and $\log z_t$. We estimate the parameters ρ_l, σ_l, and σ_z, by maximizing the Gaussian likelihood function of the vector, $X_t = (\Delta\log y_t, \log l_t)'$, subject to (24).[9] Our results are given by:

$$\log z_t = \mu_z + 0.00953\varepsilon_t^z,$$

$$\tau_{l,t} = (1-0.986)\bar{\tau}_l + 0.986\tau_{l,t-1} + 0.0056\varepsilon_t^l.$$

The *three-shock MLE specification* incorporates the investment tax shock, $\tau_{x,t}$, into the model. We estimate the three-shock MLE version of the model by maximizing the Gaussian likelihood function of the vector, $X_t = (\Delta\log y_t, \log l_t, \Delta\log i_t)'$, subject to the parameter values in (24). The results are:

$$\log z_t = \mu_z + 0.00968\varepsilon_t^z,$$

$$\tau_{l,t} = (1-0.9994)\tau_l + 0.9994\tau_{l,t-1} + 0.00631\varepsilon_t^l,$$

$$\tau_{x,t} = (1-0.9923)\tau_x + 0.9923\tau_{x,t-1} + 0.00963\varepsilon_t^x.$$

The estimated values of ρ_x and ρ_l are close to unity. This finding is consistent with other research that also reports that shocks in estimated general equilibrium models exhibit high degrees of serial correlation.[10]

2.4.2 CKM Parameterizations The *two-shock CKM specification* has two shocks, z_t and $\tau_{l,t}$. These shocks have the following time series representations:

$$\log z_t = \mu_z + 0.0131\varepsilon_t^z,$$

$$\tau_{l,t} = (1-0.952)\tau_l + 0.952\tau_{l,t-1} + 0.0136\varepsilon_t^l.$$

The *three-shock CKM specification* adds an investment shock, $\tau_{x,t}$, to the model, and has the following law of motion:

$$\tau_{x,t} = (1-0.98)\tau_x + 0.98\tau_{x,t-1} + 0.0123\varepsilon_t^x. \tag{25}$$

As in our specifications, CKM obtain their parameter estimates using maximum likelihood methods. However, their estimates are very different from ours. For example, the variances of the shocks are larger in the two-shock CKM specification than in our MLE specification. Also, the ratio of σ_l^2 to σ_z^2 is nearly three times larger in the two-shock CKM specification than in our two-shock MLE specification. Section 2.5 discusses the reasons for these differences.

2.5 *The Importance of Technology Shocks for Hours Worked*

Table 1.1 reports the contribution, V_h, of technology shocks to three different measures of the volatility in the log of hours worked: (1) the variance of the log hours, (2) the variance of HP-filtered, log hours and (3) the variance in the one-step-ahead forecast error in log hours.[11] With one exception, we compute the analogous statistics for log output. The exception is (1), for which we compute the contribution of technology shocks to the variance of the growth rate of output.

The key result in this table is that technology shocks account for a very small fraction of the volatility in hours worked. When V_h is measured according to (1), it is always below 4 percent. When V_h is measured using (2) or (3) it is always below 8 percent. For both (2) and (3), in the CKM specifications, V_h is below 2 percent.[12] Consistent with the RBC literature, the table also shows that technology accounts for a much larger movement in output.

Figure 1.1 displays visually how unimportant technology shocks are for hours worked. The top panel displays two sets of 180 artificial observations on hours worked, simulated using the standard two-shock MLE specification. The volatile time series shows how log hours worked evolve in the presence of shocks to both z_t and $\tau_{l,t}$. The other time series shows how log hours worked evolve in response to just the technology shock, z_t. The bottom panel is the analog of the top figure when the data are generated using the standard two-shock CKM specification.

3 Results Based on RBC Data Generating Mechanisms

In this section we analyze the properties of conventional VAR-based strategies for identifying the effects of a technology shock on hours worked. We focus on the bias properties of the impulse response estimator, and on standard procedures for estimating sampling uncertainty.

We use the RBC model parameterizations discussed in the previous section as the data generating processes. For each parameterization, we

Table 1.1
Contribution of Technology Shocks to Volatility

Model Specification		Unfiltered $\ln l_t$	Unfiltered $\Delta \ln y_t$	HP-Filtered $\ln l_t$	HP-Filtered $\Delta \ln y_t$	One-Step-Ahead Forecast Error $\ln l_t$	One-Step-Ahead Forecast Error $\Delta \ln y_t$
MLE							
Base	Nonrecursive	3.73	67.16	7.30	67.14	7.23	67.24
	Recursive	3.53	58.47	6.93	64.83	0.00	57.08
$\sigma_l/2$	Nonrecursive	13.40	89.13	23.97	89.17	23.77	89.16
	Recursive	12.73	84.93	22.95	88.01	0.00	84.17
$\sigma_l/4$	Nonrecursive	38.12	97.06	55.85	97.10	55.49	97.08
	Recursive	36.67	95.75	54.33	96.68	0.00	95.51
$\sigma=6$	Nonrecursive	3.26	90.67	6.64	90.70	6.59	90.61
	Recursive	3.07	89.13	6.28	90.10	0.00	88.93
$\sigma=0$	Nonrecursive	4.11	53.99	7.80	53.97	7.73	54.14
	Recursive	3.90	41.75	7.43	50.90	0.00	38.84
Three	Nonrecursive	0.18	45.67	3.15	45.69	3.10	45.72
	Recursive	0.18	36.96	3.05	43.61	0.00	39.51
CKM							
Base	Nonrecursive	2.76	33.50	1.91	33.53	1.91	33.86
	Recursive	2.61	25.77	1.81	31.41	0.00	24.93
$\sigma_l/2$	Nonrecursive	10.20	66.86	7.24	66.94	7.23	67.16
	Recursive	9.68	58.15	6.88	64.63	0.00	57.00
$\sigma_l/4$	Nonrecursive	31.20	89.00	23.81	89.08	23.76	89.08
	Recursive	29.96	84.76	22.79	87.91	0.00	84.07
$\sigma=6$	Nonrecursive	0.78	41.41	0.52	41.33	0.52	41.68
	Recursive	0.73	37.44	0.49	40.11	0.00	37.42
$\sigma=0$	Nonrecursive	2.57	20.37	1.82	20.45	1.82	20.70
	Recursive	2.44	13.53	1.73	18.59	0.00	12.33
$\sigma=0$ and $2\sigma_l$	Nonrecursive	0.66	6.01	0.46	6.03	0.46	6.12
	Recursive	0.62	3.76	0.44	5.41	0.00	3.40
Three	Nonrecursive	2.23	30.73	1.71	31.11	1.72	31.79
	Recursive	2.31	23.62	1.66	29.67	0.00	25.62

Note: (a) V_h corresponds to the columns denoted by $\ln(l_t)$.
(b) In each case, the results report the ratio of two variances:
the numerator is the variance for the system with only technology shocks and the denominator is the variance for the system with both technology shock and labor tax shocks. All statistics are averages of the ratios, based on 300 simulations of 5,000 observations for each model.
(c) "Base" means the two-shock specification, whether MLE or CKM, as indicated. "Three" means the three-shock specification.
(d) For a description of the procedure used to calculate the forecast error variance, see footnote 13.
(e) "MLE" and "CKM" refer, respectively, to our and CKM's estimated models.

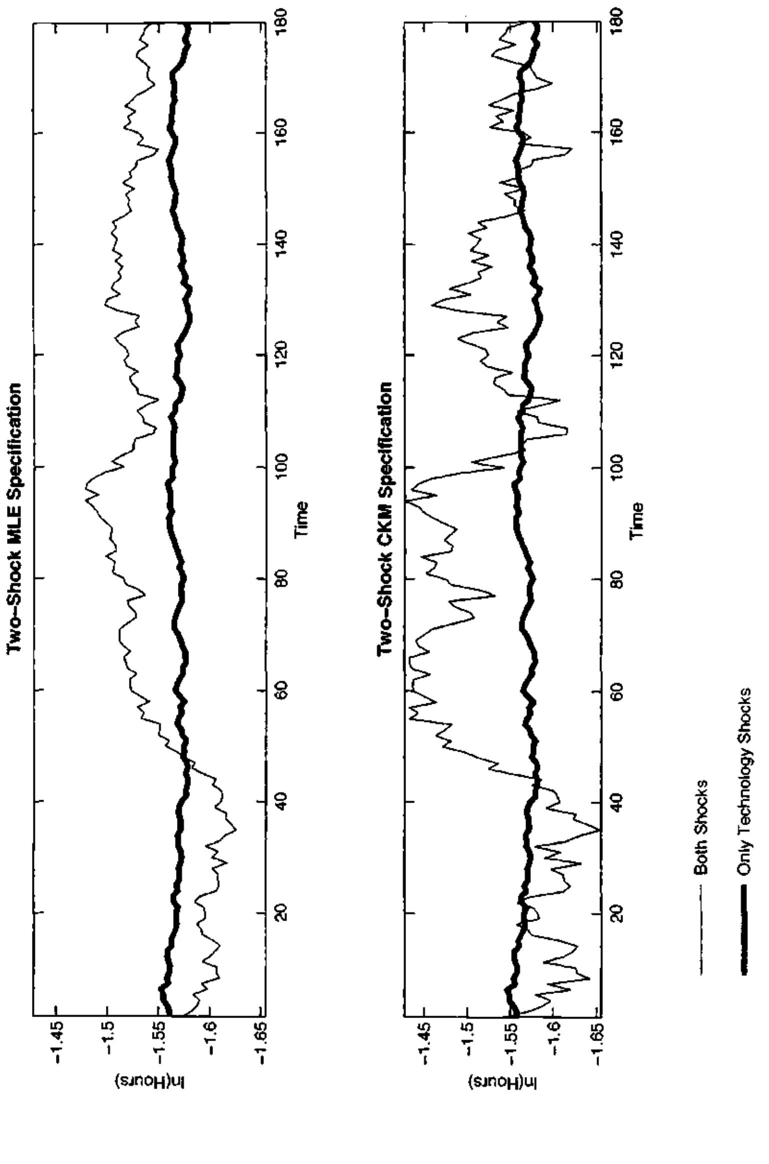


Figure 1.1
A Simulated Time Series for Hours

simulate 1,000 data sets of 180 observations each. The shocks ε^z_t, ε^l_t, and possibly ε^x_t, are drawn from i.i.d. standard normal distributions. For each artificial data set, we estimate a four-lag VAR. The average, across the 1,000 datasets, of the estimated impulse response functions, allows us to assess bias.

For each data set we also estimate two different confidence intervals: a percentile-based confidence interval and a standard-deviation based confidence interval.[13] We construct the intervals using the following bootstrap procedure. Using random draws from the fitted VAR disturbances, we use the estimated four lag VAR to generate 200 synthetic data sets, each with 180 observations. For each of these 200 synthetic data sets we estimate a new VAR and impulse response function. For each artificial data set the percentile-based confidence interval is defined as the top 2.5 percent and bottom 2.5 percent of the estimated coefficients in the dynamic response functions. The standard-deviation-based confidence interval is defined as the estimated impulse response plus or minus two standard deviations where the standard deviations are calculated across the 200 simulated estimated coefficients in the dynamic response functions.

We assess the accuracy of the confidence interval estimators in two ways. First, we compute the coverage rate for each type of confidence interval. This rate is the fraction of times, across the 1,000 data sets simulated from the economic model, that the confidence interval contains the relevant true coefficient. If the confidence intervals were perfectly accurate, the coverage rate would be 95 percent. Second, we provide an indication of the actual degree of sampling uncertainty in the VAR-based impulse response functions. In particular, we report centered 95 percent probability intervals for each lag in our impulse response function estimators.[14] If the confidence intervals were perfectly accurate, they should on average coincide with the boundary of the 95 percent probability interval.

When we generate data from the two-shock MLE and CKM specifications, we set $Y_t = (\Delta \log a_t, \log l_t)'$. When we generate data from the three-shock MLE and CKM specifications, we set $Y_t = (\Delta \log a_t, \log l_t, \log i_t/y_t)'$.

3.1 *Short-Run Identification*

Results for the two- and three- Shock MLE Specifications

Figure 1.2 reports results generated from four different parameterizations of the recursive version of the RBC model. In each panel, the

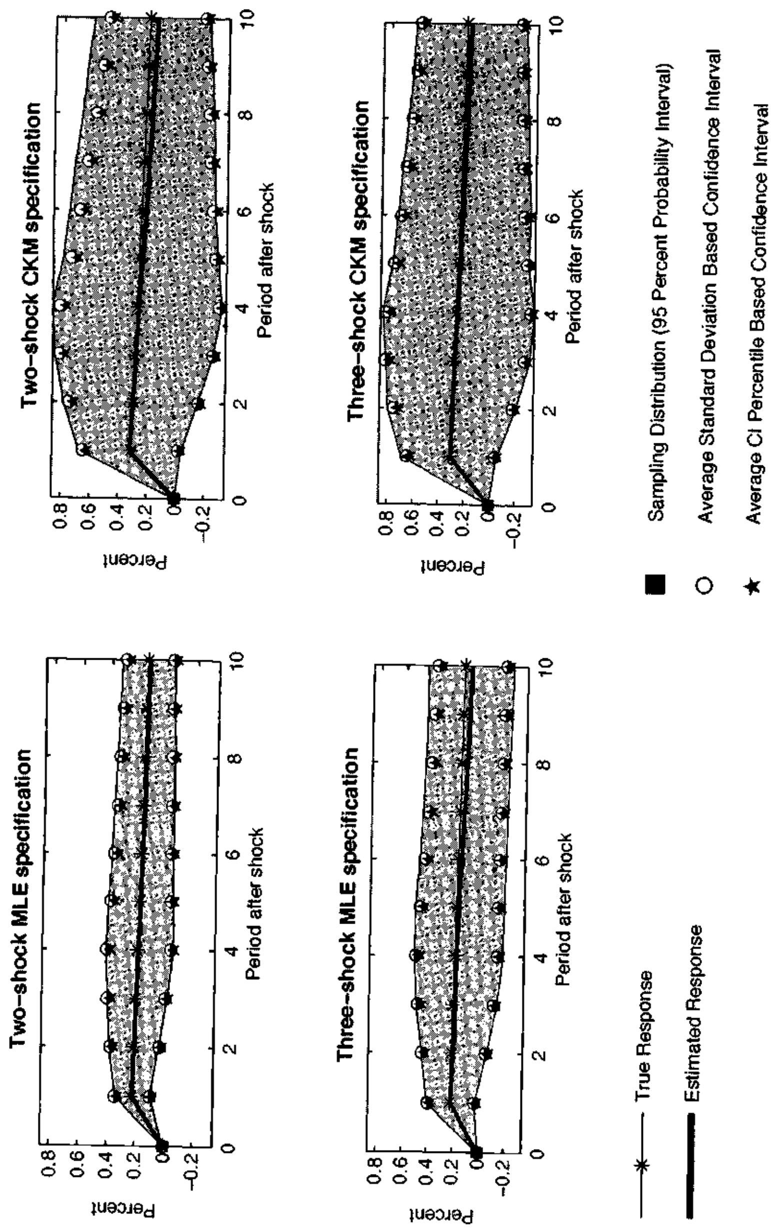


Figure 1.2
Short-Run Identification Results

solid line is the average estimated impulse response function for the 1,000 data sets simulated using the indicated economic model. For each model, the starred line is the true impulse response function of hours worked. In each panel, the gray area defines the centered 95 percent probability interval for the estimated impulse response functions. The stars with no line indicate the average percentile-based confidence intervals across the 1,000 data sets. The circles with no line indicate the average standard-deviation-based confidence intervals.

Figures 1.3 and 1.4 graph the coverage rates for the percentile-based and standard-deviation-based confidence intervals. For each case we graph how often, across the 1,000 data sets simulated from the economic model, the econometrician's confidence interval contains the relevant coefficient of the true impulse response function.

The 1,1 panel in figure 1.2 exhibits the properties of the VAR-based estimator of the response of hours to a technology shock when the data are generated by the two-shock MLE specification. The 2,1 panel corresponds to the case when the data generating process is the three-shock MLE specification.

The panels have two striking features. First, there is essentially no evidence of bias in the estimated impulse response functions. In all cases, the solid lines are very close to the starred lines. Second, an econometrician would not be misled in inference by using standard procedures for constructing confidence intervals. The circles and stars are close to the boundaries of the gray area. The 1,1 panels in figures 1.3 and 1.4 indicate that the coverage rates are roughly 90 percent. So, with high probability, VAR-based confidence intervals include the true value of the impulse response coefficients.

Results for the CKM Specification

The second column of figure 1.2 reports the results when the data generating process is given by variants of the CKM specification. The 1,2 and 2,1 panels correspond to the two and three-shock CKM specification, respectively.

The second column of figure 1.2 contains the same striking features as the first column. There is very little bias in the estimated impulse response functions. In addition, the average value of the econometrician's confidence interval coincides closely with the actual range of variation in the impulse response function (the gray area). Coverage rates, reported in the 1,2 panels of figures 1.3 and 1.4, are roughly 90

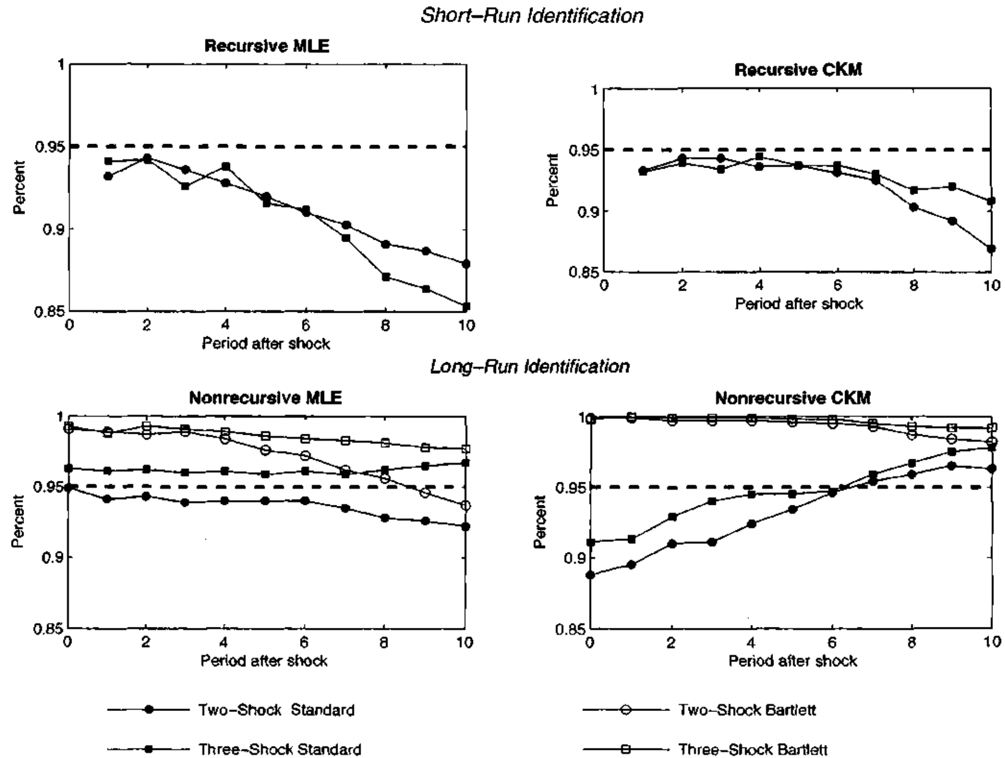


Figure 1.3
Coverage Rates for Percentile-Based Confidence Intervals

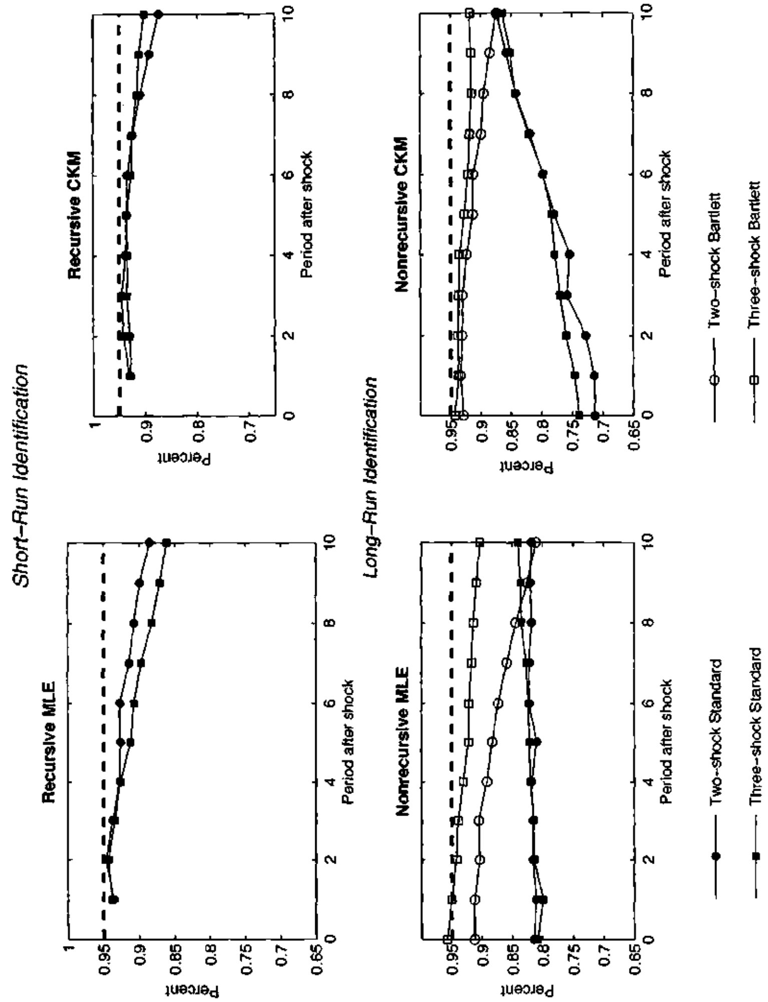


Figure 1.4
Coverage Rates for Standard Deviation-Based Confidence Intervals

percent. These rates are consistent with the view that VAR-based procedures lead to reliable inference.

A comparison of the gray areas across the first and second columns of figure 1.2, clearly indicates that more sampling uncertainty occurs when the data are generated from the CKM specifications than when they are generated from the MLE specifications (the gray areas are wider). VAR-based confidence intervals detect this fact.

3.2 *Long-run Identification*

Results for the two- and three- Shock MLE Specifications

The first and second rows of column 1 in figure 1.5 exhibit our results when the data are generated by the two- and three- shock MLE specifications. Once again there is virtually no bias in the estimated impulse response functions and inference is accurate. The coverage rates associated with the percentile-based confidence intervals are very close to 95 percent (see figure 1.3). The coverage rates for the standard-deviation-based confidence intervals are somewhat lower, roughly 80 percent (see figure 1.4). The difference in coverage rates can be seen in figure 1.5, which shows that the stars are shifted down slightly relative to the circles. Still, the circles and stars are very good indicators of the boundaries of the gray area, although not quite as good as in the analog cases in figure 1.2.

Comparing figures 1.2 and 1.5, we see that figure 1.5 reports more sampling uncertainty. That is, the gray areas are wider. Again, the crucial point is that the econometrician who computes standard confidence intervals would detect the increase in sampling uncertainty.

Results for the CKM Specification

The third and fourth rows of column 1 in figure 1.5 report results for the two- and three-shock CKM specifications. Consistent with results reported in CKM, there is substantial bias in the estimated dynamic response functions. For example, in the two-shock CKM specification, the contemporaneous response of hours worked to a one-standard-deviation technology shock is 0.3 percent, while the mean estimated response is 0.97 percent. This bias stands in contrast to our other results.

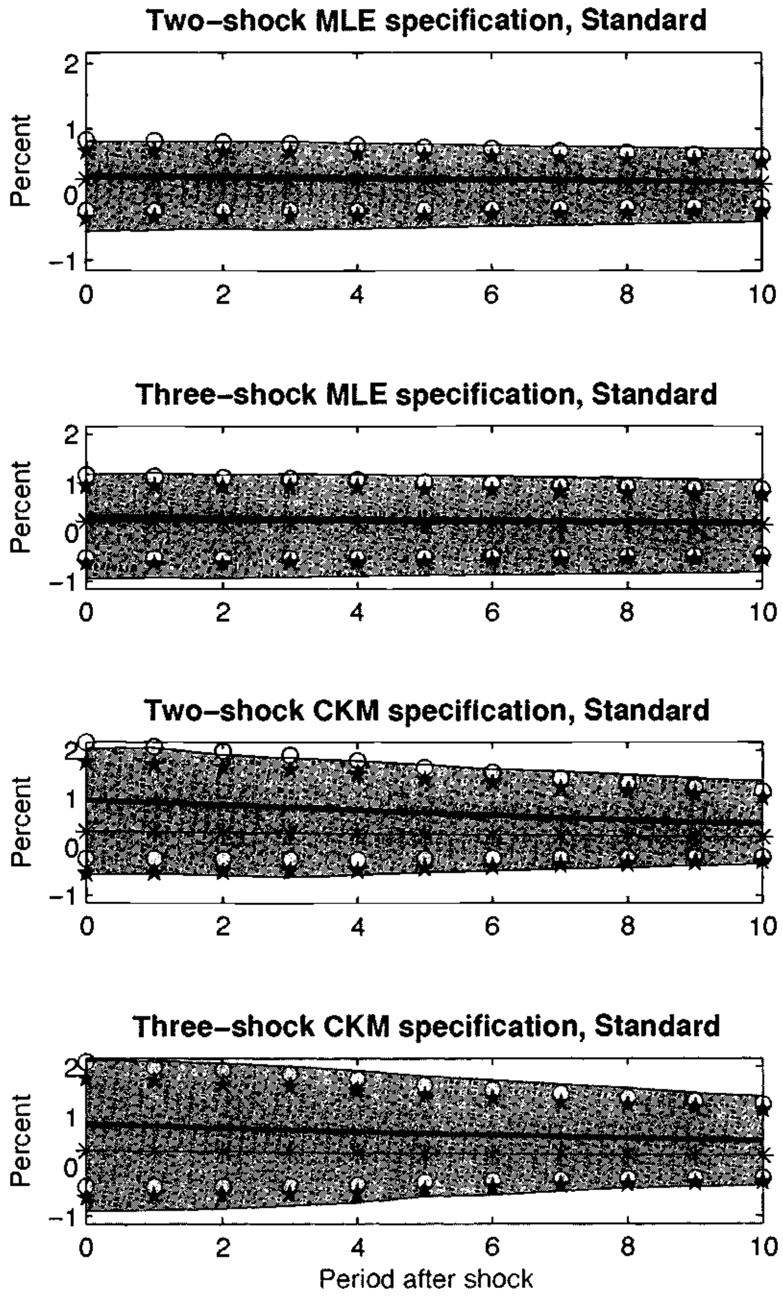


Figure 1.5
Long-Run Identification Results

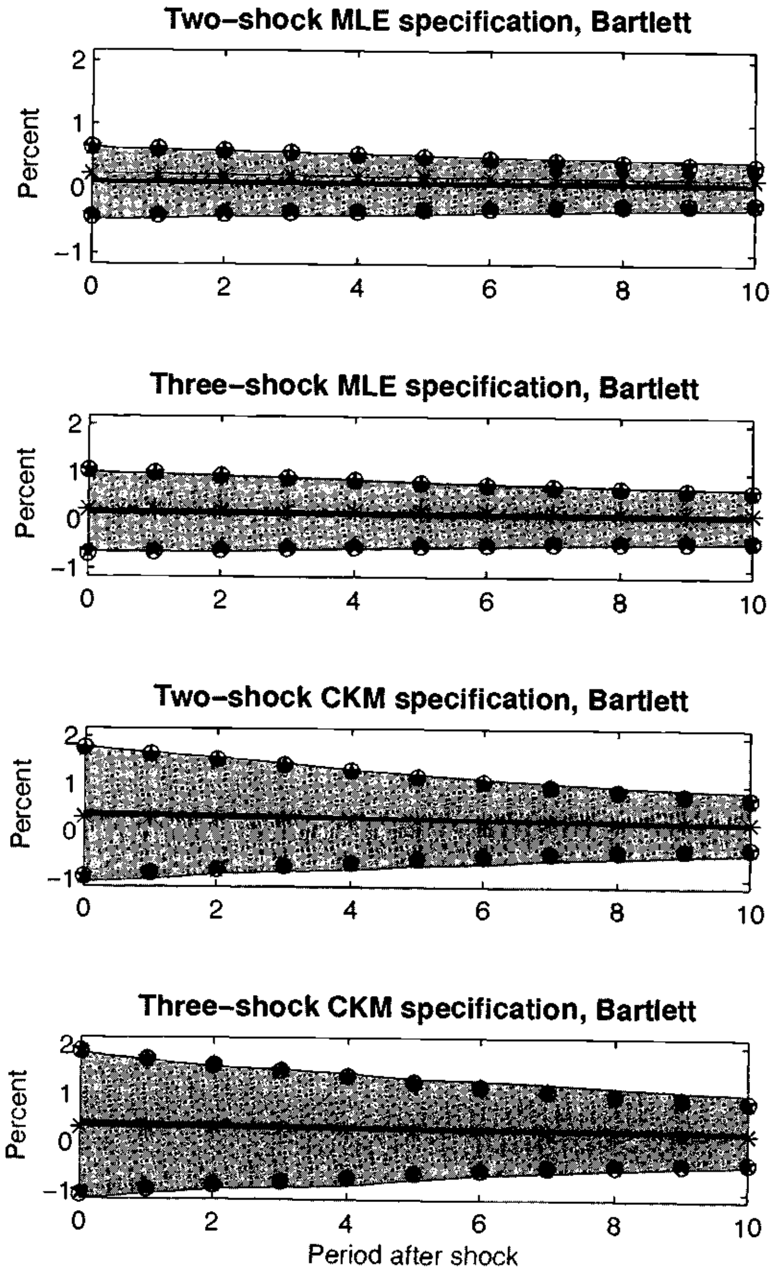


Figure 1.5 (continued)
Long-Run Identification Results

Is this bias big or problematic? In our view, bias cannot be evaluated without taking into account sampling uncertainty. Bias matters only to the extent that the econometrician is led to an incorrect inference. For example, suppose sampling uncertainty is large and the econometrician knows it. Then the econometrician would conclude that the data contain little information and, therefore, would not be misled. In this case, we say that bias is not large. In contrast, suppose sampling uncertainty is large, but the econometrician thinks it is small. Here, we would say bias is large.

We now turn to the sampling uncertainty in the CKM specifications. Figure 1.5 shows that the econometrician's average confidence interval is large relative to the bias. Interestingly, the percentile confidence intervals (stars) are shifted down slightly relative to the standard-deviation-based confidence intervals (circles). On average, the estimated impulse response function is not in the center of the percentile confidence interval. This phenomenon often occurs in practice.[15] Recall that we estimate a four lag VAR in each of our 1,000 synthetic data sets. For the purposes of the bootstrap, each of these VARs is treated as a true data generating process. The asymmetric percentile confidence intervals show that when data are generated by these VARs, VAR-based estimators of the impulse response function have a downward bias.

Figure 1.3 reveals that for the two- and three-shock CKM specifications, percentile-based coverage rates are reasonably close to 95 percent. Figure 1.4 shows that the standard deviation based coverage rates are lower than the percentile-based coverage rates. However even these coverage rates are relatively high in that they exceed 70 percent.

In summary, the results for the MLE specification differ from those of the CKM specifications in two interesting ways. First, sampling uncertainty is much larger with the CKM specification. Second, the estimated responses are somewhat biased with the CKM specification. But the bias is small: It has no substantial effect on inference, at least as judged by coverage rates for the econometrician's confidence intervals.

3.3 Confidence Intervals in the RBC Examples and a Situation in Which VAR-Based Procedures Go Awry

Here we show that the more important technology shocks are in the dynamics of hours worked, the easier it is for VARs to answer the question, "how do hours worked respond to a technology shock." We demonstrate this by considering alternative values of the innovation variance in the labor tax, σ_l, and by considering alternative values of σ, the utility parameter that controls the Frisch elasticity of labor supply.

Consider figure 1.6, which focuses on the long-run identification schemes. The first and second columns report results for the two-shock MLE and CKM specifications, respectively. For each specification we redo our experiments, reducing σ_l by a half and then by a quarter. Table 1.1 shows that the importance of technology shocks rises as the standard deviation of the labor tax shock falls. Figure 1.6 indicates that the magnitude of sampling uncertainty and the size of confidence intervals fall as the relative importance of labor tax shocks falls.[16]

Figure 1.7 presents the results of a different set of experiments based on perturbations of the two-shock CKM specification. The 1,1 and 2,1 panels show what happens when we vary the value of σ, the parameter that controls the Frisch labor supply elasticity. In the 1,1 panel we set $\sigma = 6$, which corresponds to a Frisch elasticity of 0.63. In the 2,1 panel, we set $\sigma = 0$, which corresponds to a Frisch elasticity of infinity. As the Frisch elasticity is increased, the fraction of the variance in hours worked due to technology shocks decreases (see table 1.1). The magnitude of bias and the size of confidence intervals are larger for the higher Frisch elasticity case. In both cases the bias is still smaller than the sampling uncertainty.

We were determined to construct at least one example in which the VAR-based estimator of impulse response functions has bad properties, i.e., bias is larger than sampling uncertainty. We display such an example in the 3,1 panel of figure 1.7. The data generating process is a version of the two-shock CKM model with an infinite Frisch elasticity and double the standard deviation of the labor tax rate. Table 1.1 indicates that with this specification, technology shocks account for a trivial fraction of the variance in hours worked. Of the three measures of V_h, two are 0.46 percent and the third is 0.66 percent. The 3,1 panel of figure 1.7 shows that the VAR-based procedure now has very bad properties: the true value of the impulse response function lies outside the average value of both confidence intervals that we consider. This example shows that constructing scenarios in which VAR-based procedures go awry is certainly possible. However, this example seems unlikely to be of practical significance given the poor fit to the data of this version of the model.

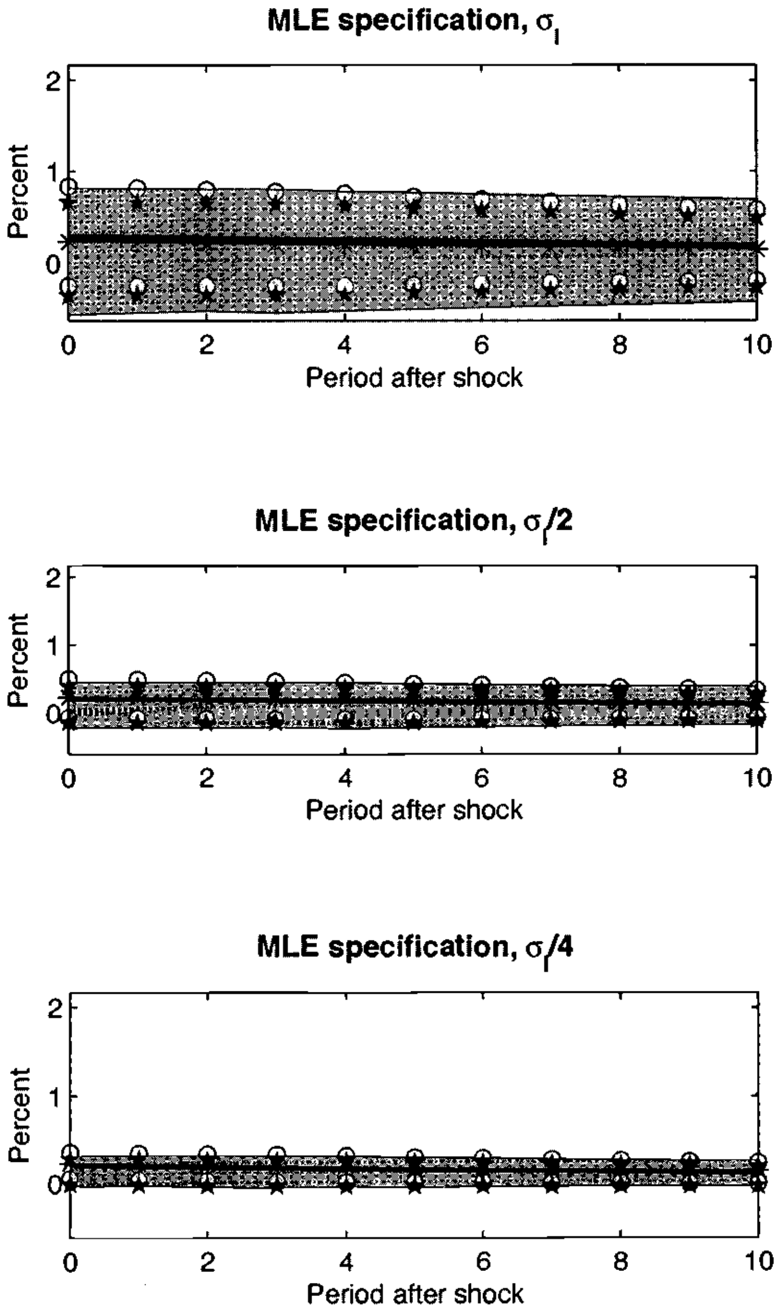


Figure 1.6
Analyzing Precision in Inference

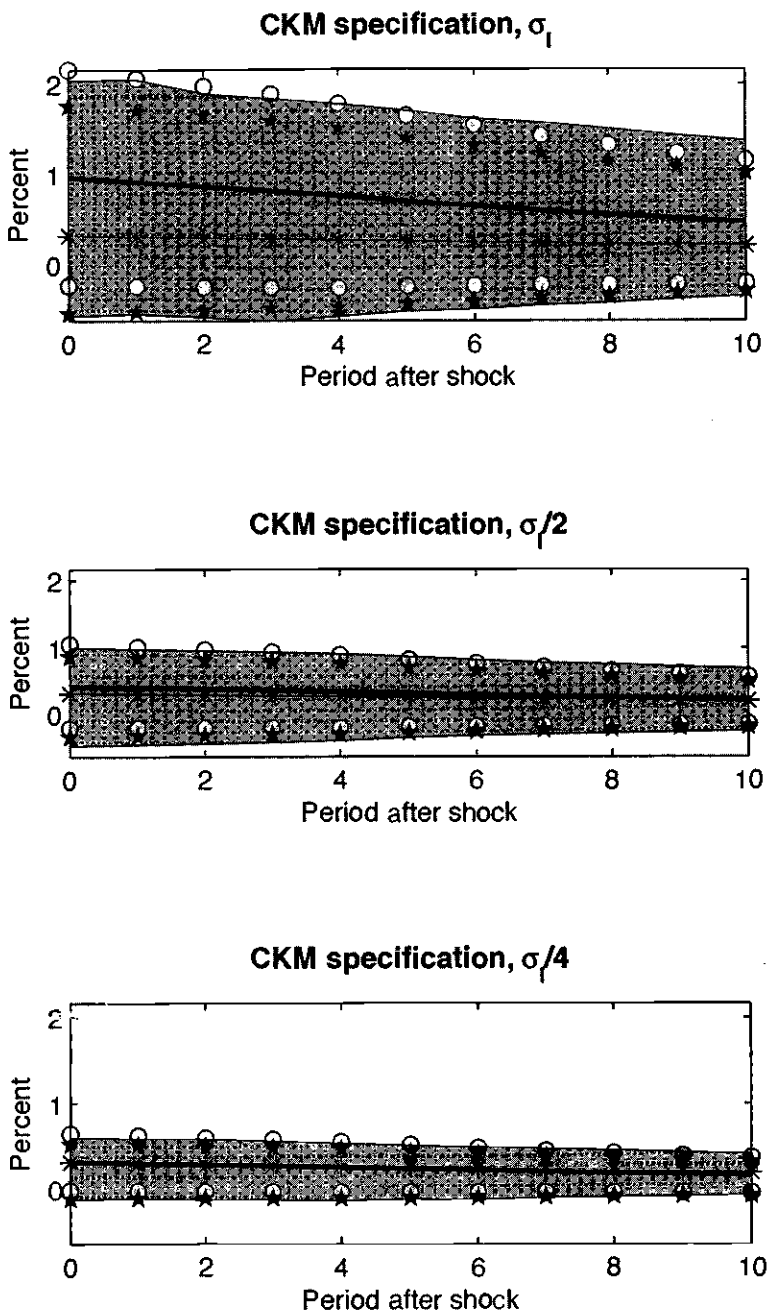


Figure 1.6 (continued)
Analyzing Precision in Inference

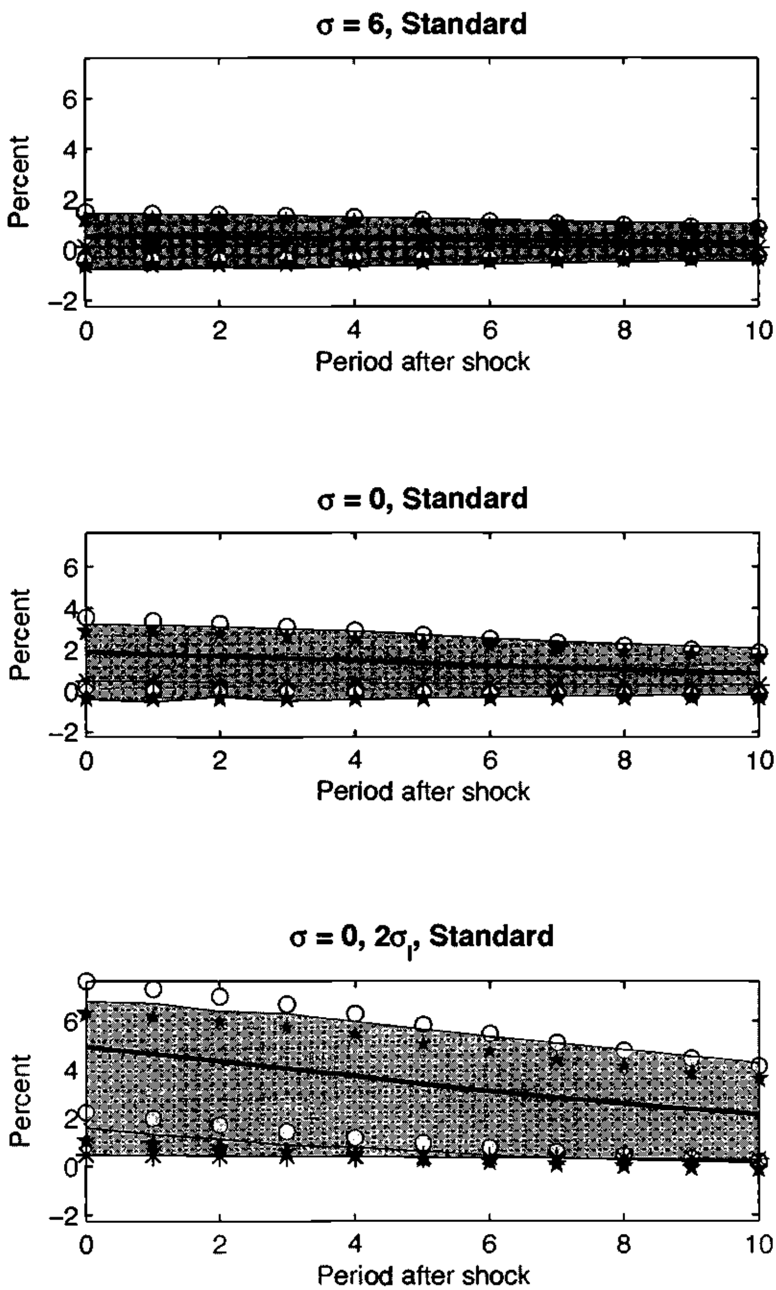


Figure 1.7
Varying the Labor Elasticity in the Two-Shock CKM Specification

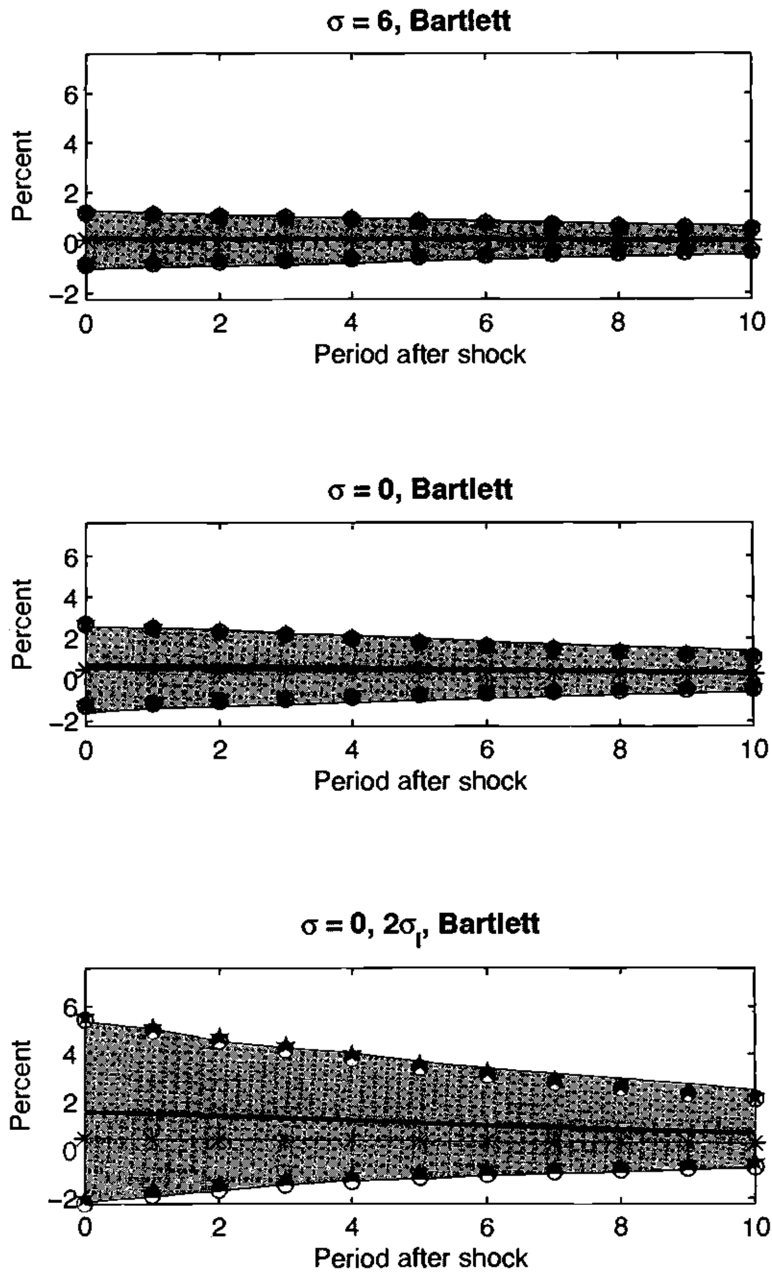


Figure 1.7 (continued)
Varying the Labor Elasticity in the Two-Shock CKM Specification

3.4 Are Long-Run Identification Schemes Informative?

Up to now, we have focused on the RBC model as the data generating process. For empirically reasonable specifications of the RBC model, confidence intervals associated with long-run identification schemes are large. One might be tempted to conclude that VAR-based long-run identification schemes are uninformative. Specifically, are the confidence intervals so large that we can never discriminate between competing economic models? Erceg, Guerrieri, and Gust (2005) show that the answer to this question is "no." They consider an RBC model similar to the one discussed above and a version of the sticky wage-price model developed by Christiano, Eichenbaum, and Evans (2005) in which hours worked fall after a positive technology shock. They then conduct a series of experiments to assess the ability of a long-run identified structural VAR to discriminate between the two models on the basis of the response of hours worked to a technology shock.

Using estimated versions of each of the economic models as a data generating process, they generate 10,000 synthetic data sets each with 180 observations. They then estimate a four-variable structural VAR on each synthetic data set and compute the dynamic response of hours worked to a technology shock using long-run identification. Erceg, Guerrieri, and Gust (2005) report that the probability of finding an initial decline in hours that persists for two quarters is much higher in the model with nominal rigidities than in the RBC model (93 percent versus 26 percent). So, if these are the only two models contemplated by the researcher, an empirical finding that hours worked decline after a positive innovation to technology will constitute compelling evidence in favor of the sticky wage-price model.

Erceg, Guerrieri, and Gust (2005) also report that the probability of finding an initial rise in hours that persists for two quarters is much higher in the RBC model than in the sticky wage-price model (71 percent versus 1 percent). So, an empirical finding that hours worked rises after a positive innovation to technology would constitute compelling evidence in favor of the RBC model versus the sticky wage-price alternative.

4 Contrasting Short- and Long-Run Restrictions

The previous section demonstrates that, in the examples we considered, when VARs are identified using short-run restrictions, the conventional estimator of impulse response functions is remarkably accurate. In contrast, for some parameterizations of the data generating process, the conventional estimator of impulse response functions based on long-run identifying restrictions can exhibit noticeable bias. In this section we argue that the key difference between the two identification strategies is that the long-run strategy requires an estimate of the sum of the VAR coefficients, $B(1)$. This object is notoriously difficult to estimate accurately (see Sims 1972).

We consider a simple analytic expression related to one in Sims (1972). Our expression shows what an econometrician who fits a misspecified, fixed-lag, finite-order VAR would find in population. Let $\hat{B}_1, \ldots, \hat{B}_q$, and $\hat{V}$ denote the parameters of the qth-order VAR fit by the econometrician. Then:

$$\hat{V} = V + \min_{\hat{B}_1,\ldots\hat{B}_q} \frac{1}{2\pi}\int_{-\pi}^{\pi} [B(e^{-i\omega}) - \hat{B}(e^{-i\omega})] S_Y(\omega) [B(e^{i\omega}) - \hat{B}(e^{i\omega})]' d\omega, \tag{26}$$

where

$$B(L) = B_1 + B_2 L + B_3 L^2 + \ldots,$$

$$\hat{B}(L) = \hat{B}_1 + \hat{B}_2 L + \ldots + \hat{B}_4 L^3.$$

Here, $B(e^{-i\omega})$ and $\hat{B}(e^{-i\omega})$ correspond to $B(L)$ and $\hat{B}(L)$ with L replaced by $e^{-i\omega}$.[17] In (26), B and V are the parameters of the actual infinite-ordered VAR representation of the data (see (10)), and $S_Y(\omega)$ is the associated spectral density at frequency ω.[18] According to (26), estimation of a VAR approximately involves choosing VAR lag matrices to minimize a quadratic form in the difference between the estimated and true lag matrices. The quadratic form assigns greatest weight to the frequencies for which the spectral density is the greatest. If the econometrician's VAR is correctly specified, then $\hat{B}(e^{-i\omega}) = B(e^{-i\omega})$ for all ω, and $\hat{V} = V$, so that the estimator is consistent. If there is specification error, then $\hat{B}(e^{-i\omega}) \neq B(e^{-i\omega})$ for some ω and $V > \hat{V}$.[19] In our context, specification error exists because the true VAR implied by our data generating processes has $q = \infty$, but the econometrician uses a finite value of q.

To understand the implications of (26) for our analysis, it is useful to write in lag-operator form the estimated dynamic response of Y_t to a shock in the first element of ε_t

$$Y_t = [I + \theta_1 L + \theta_2 L^2 + \ldots]\hat{C}_1 \varepsilon_{1,t}, \tag{27}$$

where the θ_k's are related to the estimated VAR coefficients as follows:

$$\theta_k = \frac{1}{2\pi}\int_{-\pi}^{\pi}[I - \hat{B}(e^{-i\omega})e^{-i\omega}]^{-1}e^{k\omega i}d\omega. \tag{28}$$

In the case of long-run identification, the vector $\hat{C}_1$ is computed using (22), and $\hat{B}(1)$ and $\hat{V}$ replace $B(1)$ and V respectively. In the case of short-run identification, we compute $\hat{C}_1$ as the second column in the upper triangular Cholesky decomposition of $\hat{V}$.[20]

We use (26) to understand why estimation based on short-run and long-run identification can produce different results. According to (27), impulse response functions can be decomposed into two parts, the impact effect of the shocks, summarized by $\hat{C}_1$, and the dynamic part summarized in the term in square brackets. We argue that when a bias arises with long-run restrictions, it is because of difficulties in estimating C_1. These difficulties do not arise with short-run restrictions.

In the short-run identification case, $\hat{C}_1$ is a function of $\hat{V}$ only. Across a variety of numerical examples, we find that $\hat{V}$ is very close to V.[21] This result is not surprising because (26) indicates that the entire objective of estimation is to minimize the distance between $\hat{V}$ and V. In the long-run identification case, $\hat{C}_1$ depends not only on $\hat{V}$ but also on $\hat{B}(1)$. A problem is that the criterion does not assign much weight to setting $\hat{B}(1) = B(1)$ unless $S_Y(\omega)$ happens to be relatively large in a neighborhood of $\omega = 0$. But, a large value of $S_Y(0)$ is not something one can rely on.[22] When $S_Y(0)$ is relatively small, attempts to match $\hat{B}(e^{-i\omega})$ with $B(e^{-i\omega})$ at other frequencies can induce large errors in $\hat{B}(1)$.

The previous argument about the difficulty of estimating C_1 in the long-run identification case does not apply to the θ'_ks. According to (28) θ_k is a function of $\hat{B}(e^{-i\omega})$ over the whole range of ω's, not just one specific frequency.

We now present a numerical example, which illustrates Proposition 1 as well as some of the observations we have made in discussing (26). Our numerical example focuses on population results. Therefore, it provides only an indication of what happens in small samples.

To understand what happens in small samples, we consider four additional numerical examples. First, we show that when the econometrician uses the true value of $B(1)$, the bias and much of the sampling uncertainty associated with the two-shock CKM specification disappears. Second, we demonstrate that bias problems essentially disappear when we use an alternative to the standard zero-frequency spectral density estimator used in the VAR literature. Third, we show that the problems are attenuated when the preference shock is more persistent. Fourth, we consider the recursive version of the two-shock CKM specification in which the effect of technology shocks can be estimated using either short- or long-run restrictions.

A Numerical Example

Table 1.2 reports various properties of the two-shock CKM specification. The first six B_j's in the infinite-order VAR, computed using (12), are reported in Panel A. These B_j's eventually converge to zero, however they do so slowly. The speed of convergence is governed by the size of the maximal eigenvalue of the matrix M in (8), which is 0.957. Panel B displays the $\hat{B}_j$'s that solve (26) with $q = 4$. Informally, the $\hat{B}_j$'s look similar to the B_j's for $j = 1, 2, 3, 4$. In line with this observation, the sum of the true B_j's, $B_1 + \ldots + B_4$ is similar in magnitude to the sum of the estimated $\hat{B}_j$'s, $\hat{B}(1)$ (see Panel C). But the econometrician using long-run restrictions needs a good estimate of $B(1)$. This matrix is very different from $B_1 + \ldots + B_4$. Although the remaining B_j's for $j > 4$ are individually small, their sum is not. For example, the 1,1 element of $B(1)$ is 0.28, or six times larger than the 1,1 element of $B_1 + \ldots + B_4$.

The distortion in $\hat{B}(1)$ manifests itself in a distortion in the estimated zero-frequency spectral density (see Panel D). As a result, there is distortion in the estimated impact vector, $\hat{C}_1$ (Panel F).[23] To illustrate the significance of the latter distortion for estimated impulse response functions, we display in figure 1.8 the part of (27) that corresponds to the response of hours worked to a technology shock. In addition, we display the true response. There is a substantial distortion, which is approximately the same magnitude as the one reported for small samples in figure 1.5. The third line in figure 1.8 corresponds to (27) when $\hat{C}_1$ is replaced by its true value, C_1. Most of the distortion in the estimated impulse response function is eliminated by this replacement. Finally, the distortion in $\hat{C}_1$ is due to distortion in $\hat{B}(1)$, as $\hat{V}$ is virtually identical to V (Panel E).

This example is consistent with our overall conclusion that the individual B_j's and V are well estimated by the econometrician using a four-lag VAR. The distortions that arise in practice primarily reflect difficulties in estimating $B(1)$. Our short-run identification results in figure 1.2 are consistent with this claim, because distortions are minimal with short-run identification.

Table 1.2
Properties of Two-Shock CKM Specification

Panel A: First Six Lag Matrices in Infinite-Order VAR Representation

$$B_1 = \begin{bmatrix} 0.013 & 0.041 \\ 0.0065 & 0.94 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0.012 & -0.00 \\ 0.0062 & -0.00 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 0.012 & -0.00 \\ 0.0059 & -0.00 \end{bmatrix},$$

$$B_4 = \begin{bmatrix} 0.011 & -0.00 \\ 0.0056 & -0.00 \end{bmatrix}, \quad B_5 = \begin{bmatrix} 0.011 & -0.00 \\ 0.0054 & -0.00 \end{bmatrix}, \quad B_6 = \begin{bmatrix} 0.010 & -0.00 \\ 0.0051 & -0.00 \end{bmatrix}$$

Panel B: Population Estimate of Four-lag VAR

$$\hat{B}_1 = \begin{bmatrix} 0.017 & 0.043 \\ 0.0087 & 0.94 \end{bmatrix}, \quad \hat{B}_2 = \begin{bmatrix} 0.017 & -0.00 \\ 0.0085 & -0.00 \end{bmatrix}, \quad \hat{B}_3 = \begin{bmatrix} 0.012 & -0.00 \\ 0.0059 & -0.00 \end{bmatrix},$$

$$\hat{B}_4 = \begin{bmatrix} 0.0048 & -0.0088 \\ 0.0025 & -0.0045 \end{bmatrix}$$

Panel C: Actual and Estimated Sum of VAR Coefficients

$$\hat{B}(1) = \begin{bmatrix} 0.055 & 0.032 \\ 0.14 & 0.94 \end{bmatrix}, \quad B(1) = \begin{bmatrix} 0.28 & 0.022 \\ 0.14 & 0.93 \end{bmatrix}, \quad \Sigma_{j=1}^{4} B_j = \begin{bmatrix} 0.047 & 0.039 \\ 0.024 & 0.94 \end{bmatrix}$$

Panel D: Actual and Estimated Zero-Frequency Spectral Density

$$S_Y(0) = \begin{bmatrix} 0.00017 & 0.00097 \\ 0.00097 & 0.12 \end{bmatrix}, \quad \hat{S}_Y(0) = \begin{bmatrix} 0.00012 & 0.0022 \\ 0.0022 & 0.13 \end{bmatrix}.$$

Panel E: Actual and Estimated One-Step-Ahead Forecast Error Variance

$$V = \hat{V} = \begin{bmatrix} 0.00012 & -0.00015 \\ -0.00015 & -0.00053 \end{bmatrix}$$

Panel F: Actual and Estimated Impact Vector

$$C_1 = \begin{pmatrix} 0.00773 \\ 0.00317 \end{pmatrix}, \quad \hat{C}_1 = \begin{pmatrix} 0.00406 \\ 0.01208 \end{pmatrix}$$

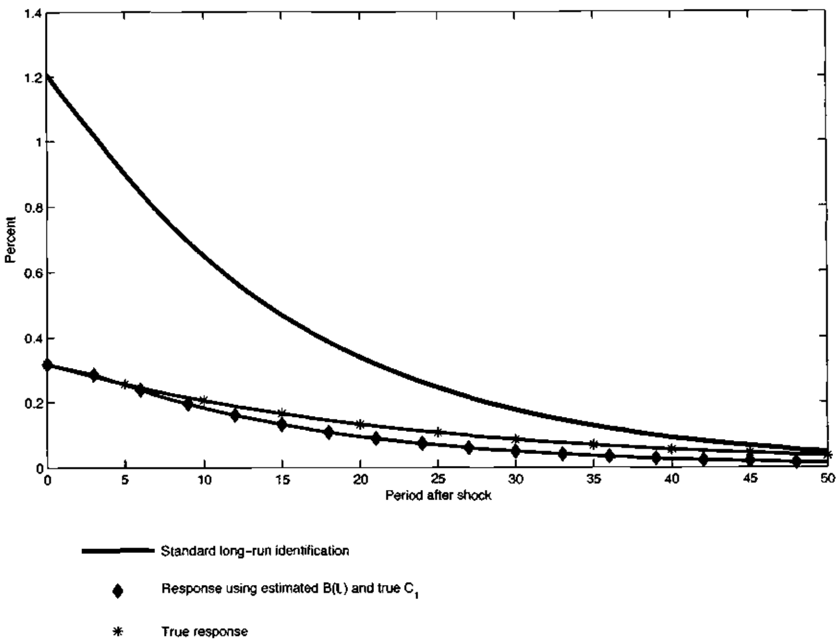


Figure 1.8
Effect of C_1 on Distortions

Using the True Value of B(1) in a Small Sample

A natural way to isolate the role of distortions in $\hat{B}(1)$ is to replace $\hat{B}(1)$ by its true value when estimating the effects of a technology shock. We perform this replacement for the two-shock CKM specification, and report the results in figure 1.9. For convenience, the 1,1 panel of figure 1.9 repeats our results for the two-shock CKM specification from the 3,1 panel in figure 1.5. The 1,2 panel of figure 1.9 shows the sampling properties of our estimator when the true value of $B(1)$ is used in repeated samples. When we use the true value of $B(1)$ the bias completely disappears. In addition, coverage rates are much closer to 95 percent and the boundaries of the average confidence intervals are very close to the boundaries of the gray area.

Using an Alternative Zero-Frequency Spectral Density Estimator

In practice, the econometrician does not know $B(1)$. However, we can replace the VAR-based zero-frequency spectral density in (19) with an

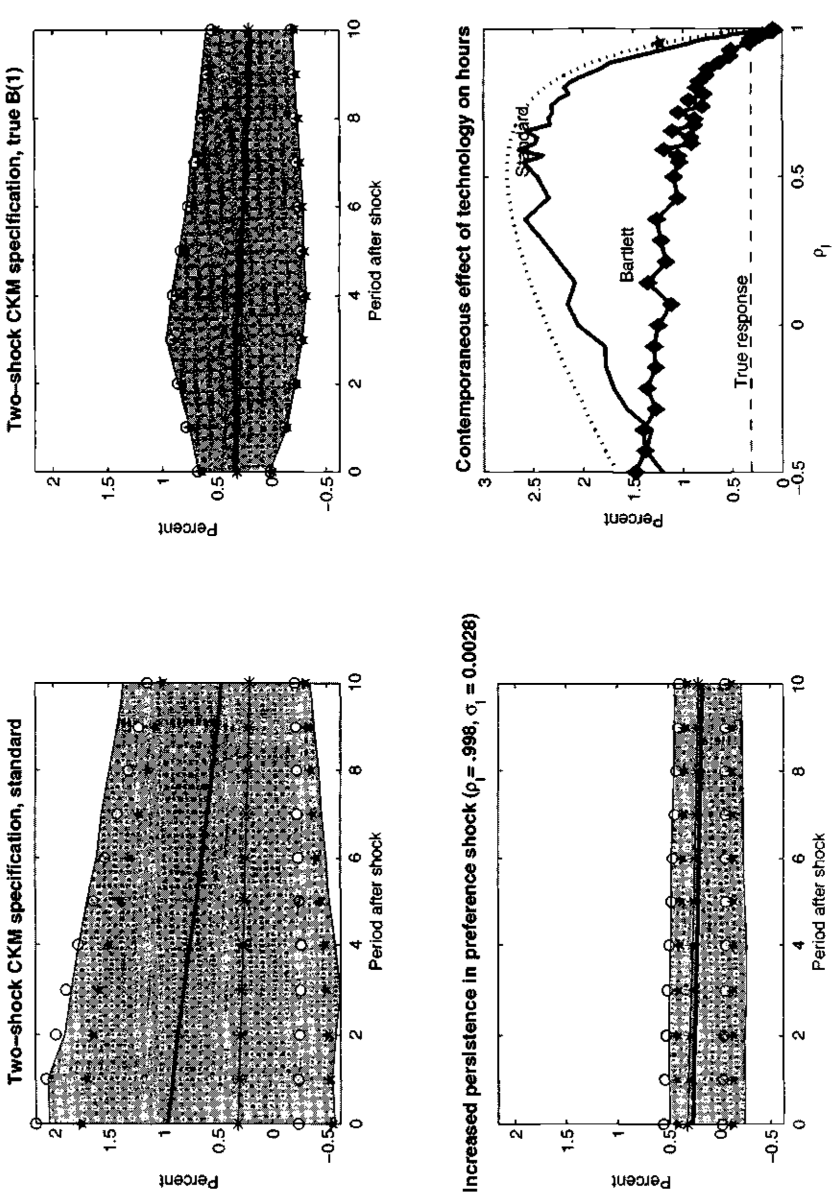


Figure 1.9
Analysis of Long-Run Identification Results

alternative estimator of $S_Y(0)$. Here, we consider the effects of using a standard Bartlett estimator:[24]

$$S_Y(0) = \sum_{k=-(T-1)}^{T-1} g(k)\hat{C}(k), \quad g(k) = \begin{cases} 1 - \frac{|k|}{r} & |k| \leq r \\ 0 & |k| > r \end{cases}, \tag{29}$$

where, after removing the sample mean from Y_t:

$$\hat{C}(k) = \frac{1}{T} \sum_{t=k+1}^{T} Y_t Y'_{t-k}.$$

We use essentially all possible covariances in the data by choosing a large value of r, $r = 150$.[25] In some respects, our modified estimator is equivalent to running a VAR with longer lags.

We now assess the effect of our modified long-run estimator. The first two rows in figure 1.5 present results for cases in which the data generating mechanism corresponds to our two-and three-shock MLE specifications. Both the standard estimator (the left column) and our modified estimator (the right column) exhibit little bias. In the case of the standard estimator, the econometrician's estimator of standard errors understates somewhat the degree of sampling uncertainty associated with the impulse response functions. The modified estimator reduces this discrepancy. Specifically, the circles and stars in the right column of figure 1.5 coincide closely with the boundary of the gray area. Coverage rates are reported in the 2,1 panels of figures 1.3 and 1.4. In figure 1.3, coverage rates now exceed 95 percent. The coverage rates in figure 1.4 are much improved relative to the standard case. Indeed, these rates are now close to 95 percent. Significantly, the degree of sampling uncertainty associated with the modified estimator is not greater than that associated with the standard estimator. In fact, in some cases, sampling uncertainty declines slightly.

The last two rows of column 1 in figure 1.5 display the results when the data generating process is a version of the CKM specification. As shown in the second column, the bias is essentially eliminated by using the modified estimator. Once again the circles and stars roughly coincide with the boundary of the gray area. Coverage rates for the percentile-based confidence intervals reported in figure 1.3 again have a tendency to exceed 95 percent (2,2 panel). As shown in the 2,2 panel of figure 1.4, coverage rates associated with the standard deviation based estimator are very close to 95 percent. There is a substantial improve-ment over the coverage rates associated with the standard spectral density estimator.

Figure 1.5 indicates that when the standard estimator works well, the modified estimator also works well. When the standard estimator results in biases, the modified estimator removes them. These findings are consistent with the notion that the biases for the two CKM specifications reflect difficulties in estimating the spectral density at frequency zero. Given our finding that $\hat{V}$ is an accurate estimator of V, we conclude that the difficulties in estimating the zero-frequency spectral density in fact reflect problems with $B(1)$.

The second column of figure 1.7 shows how our modified VAR-based estimator works when the data are generated by the various perturbations on the two-shock CKM specification. In every case, bias is substantially reduced.

Shifting Power to the Low Frequencies

Formula (26), suggests that, other things being equal, the more power there is near frequency zero, the less bias there is in $\hat{B}(1)$ and the better behaved is the estimated impulse response function to a technology shock. To pursue this observation we change the parameterization of the non-technology shock in the two-shock CKM specification. We reallocate power toward frequency zero, holding the variance of the shock constant by increasing ρ_l to 0.998 and suitably lowering σ_l in (1). The results are reported in the 2,1 panel of figure 1.9. The bias associated with the two-shock CKM specification almost completely disappears. This result is consistent with the notion that the bias problems with the two-shock CKM specification stem from difficulties in estimating $B(1)$.

The previous result calls into question conjectures in the literature (see Erceg, Guerrieri, and Gust 2005). According to these conjectures, if there is more persistence in a non-technology shock, then the VAR will produce biased results because it will confuse the technology and non-technology shocks. Our result shows that this intuition is incomplete, because it fails to take into account all of the factors mentioned in our discussion of (26). To show the effect of persistence, we consider a range of values of ρ_l to show that the impact of ρ_l on bias is in fact not monotone.

The 2,2 panel of figure 1.9 displays the econometrician's estimator of the contemporaneous impact on hours worked of a technology shock against ρ_l. The dashed line indicates the true contemporaneous effect of a technology shock on hours worked in the two-shock CKM specification. The dot-dashed line in the figure corresponds to the solution of (26), with $q = 4$, using the standard VAR-based estimator.[26] The star in the figure indicates the value of ρ_l in the two-shock CKM specification. In the neighborhood of this value of ρ_l, the distortion in the estimator falls sharply as ρ_l increases. Indeed, for $\rho_l = 0.9999$, essentially no distortion occurs. For values of ρ_l in the region, (–0.5, 0.5), the distortion increases with increases in ρ_l.

The 2,2 panel of figure 1.9 also allows us to assess the value of our proposed modification to the standard estimator. The line with diamonds displays the modified estimator of the contemporaneous impact on hours worked of a technology shock. When the standard estimator works well, that is, for large values of ρ_l the modified and standard estimators produce similar results. However, when the standard estimator works poorly, e.g., for values of ρ_l near 0.5, our modified estimator cuts the bias in half.

A potential shortcoming of the previous experiments is that persistent changes in $\tau_{l,t}$ do not necessarily induce very persistent changes in labor productivity. To assess the robustness of our results, we also considered what happens when there are persistent changes in $\tau_{x,t}$. These do have a persistent impact on labor productivity. In the two-shock CKM model, we set $\tau_{l,t}$ to a constant and allowed $\tau_{x,t}$ to be stochastic. We considered values of ρ_x in the range, [–0.5, 1], holding the variance of $\tau_{x,t}$ constant. We obtain results similar to those reported in the 2,2 panel of figure 1.9.

Short- and Long-Run Restrictions in a Recursive Model

We conclude this section by considering the recursive version of the two-shock CKM specification. This specification rationalizes estimating the impact on hours worked of a shock to technology using either the short- or the long-run identification strategy. We generate 1,000 data sets, each of length 180. On each synthetic data set, we estimate a four lag, bivariate VAR. Given this estimated VAR, we can estimate the effect of a technology shock using the short- and long-run identification strategy. Figure 1.10 reports our results. For the long-run identification strategy, there is substantial bias. In sharp contrast, there is no bias for the short-run identification strategy. Because both procedures use the same estimated VAR parameters, the bias in the long-run identification strategy is entirely attributable due to the use of $\hat{B}(1)$.

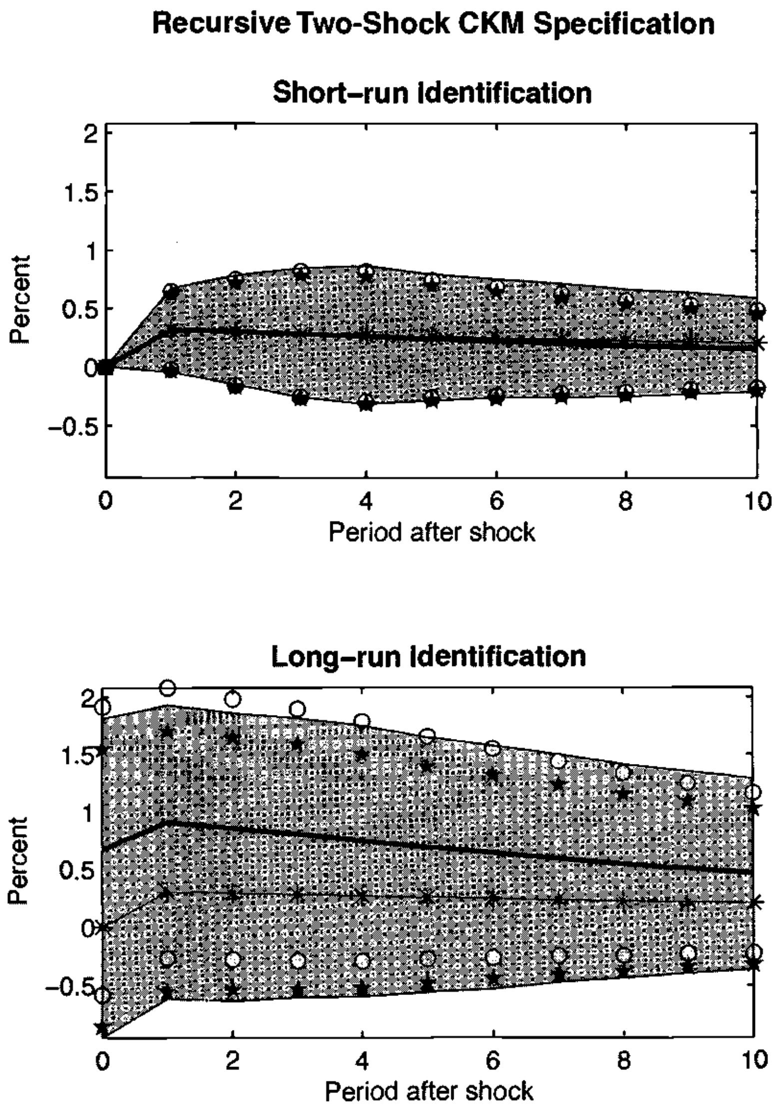


Figure 1.10
Comparing Long- and Short-Run Identifications

5 Relation to Chari-Kehoe-McGrattan

In the preceding sections we argue that structural VAR-based procedures have good statistical properties. Our conclusions about the usefulness of structural VARs stand in sharp contrast to the conclusions of CKM. These authors argue that, for plausibly parameterized RBC models, structural VARs lead to misleading results. They conclude that structural VARs are not useful for constructing and evaluating structural economic models. In this section we present the reasons we disagree with CKM.

CKM's Exotic Data Generating Processes

CKM's critique of VARs is based on simulations using particular DSGE models estimated by maximum likelihood methods. Here, we argue that their key results are driven by assumptions about measurement error. CKM's measurement error assumptions are overwhelmingly rejected in favor of alternatives under which their key results are overturned.

CKM adopt a state-observer setup to estimate their model. Define:

$$Y_t = (\Delta \log a_t, \log l_t, \Delta \log i_t, \Delta \log G_t)',$$

where G_t denotes government spending plus net exports. CKM suppose that

$$Y_t = X_t + v_t, \quad E v_t v_t' = R, \tag{30}$$

where R is diagonal, v_t is a 4×1 vector of i.i.d. measurement errors and X_t is a 4×1 vector containing the model's implications for the variables in Y_t. The two-shock CKM specification has only the shocks, $\tau_{l,t}$ and z_t. CKM model government spending plus net exports as:

$$G_t = g_t \times Z_t,$$

where g_t is in principle an exogenous stochastic process. However, when CKM estimate the parameters of the technology and preferences processes, $\tau_{l,t}$ and z_t, they set the variance of the government spending shock to zero, so that g_t is a constant. As a result, CKM assume that

$$\Delta \log G_t = \log z_t + \text{measurement error}.$$

CKM fix the elements on the diagonal of R exogenously to a "small number," leading to the remarkable implication that government purchases plus net exports.

To demonstrate the sensitivity of CKM's results to their specification of the magnitude of R, we consider the different assumptions that CKM make in different drafts of their paper. In the draft of May 2005, CKM (2005a) set the diagonal elements of R to 0.0001. In the draft of July 2005, CKM (2005b) set the i^{th} diagonal element of R equal to 0.01 times the variance of the i^{th} element of Y_t.

The 1,1 and 2,1 panels in figure 1.11 report results corresponding to CKM's two-shock specifications in the July and May drafts, respectively.[27] These panels display the log likelihood value (*see LLF*) of these two models and their implications for VAR-based impulse response functions (the 1,1 panel is the same as the 3,1 panel in figure 1.5). Surprisingly, the log-likelihood of the July specification is orders of magnitude worse than that of the May specification.

The 3,1 panel in figure 1.11 displays our results when the diagonal elements of R are included among the parameters being estimated.[28] We refer to the resulting specification as the "CKM free measurement error specification." First, both the May and the July specifications are rejected relative to the free measurement error specification. The likelihood ratio statistic for testing the May and July specifications are 428 and 6,266, respectively. Under the null hypothesis that the May or July specification is true, these statistics are realizations of a chi-square distribution with four degrees of freedom. The evidence against CKM's May or July specifications of measurement error is overwhelming.

Second, when the data generating process is the CKM free measurement error specification, the VAR-based impulse response function is virtually unbiased (see the 3,1 panel in figure 1.11). We conclude that the bias in the two-shock CKM specification is a direct consequence of CKM's choice of the measurement error variance.

As noted above, CKM's measurement error assumption has the implication that $\Delta \log G_t$ is roughly equals to $\log z_t$. To investigate the role played by this peculiar implication, we delete $\Delta \log G_t$ from Y_t and re-estimate the system. We present the results in the right column of figure 1.11. In each panel of that column, we re-estimate the system in the same way as the corresponding panel in the left column, except that $\Delta \log G_t$ is excluded from Y_t. Comparing the 2,1 and 2,2 panels, we see that, with the May measurement error specification, the bias disappears after relaxing CKM's $\Delta \log G_t = \log z_t$ assumption. Under the July

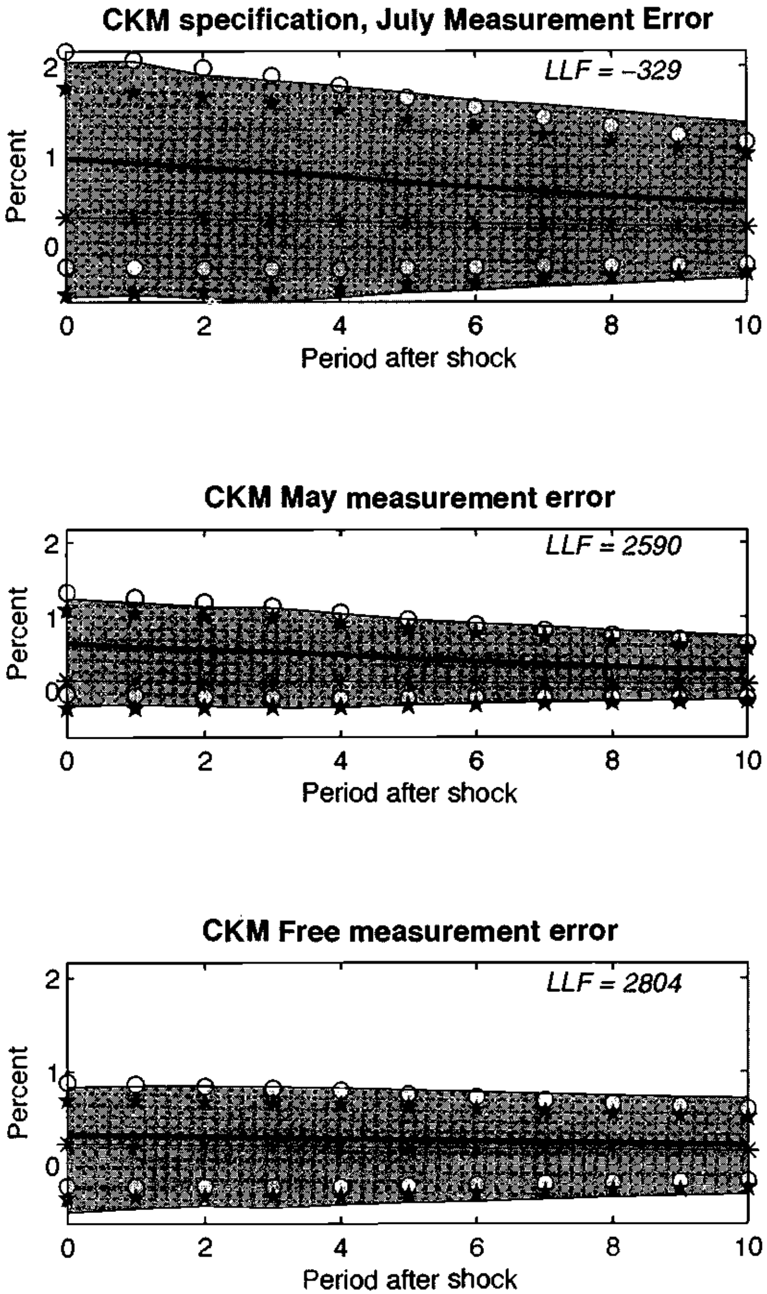


Figure 1.11
The Treatment of CKM Measurement Error

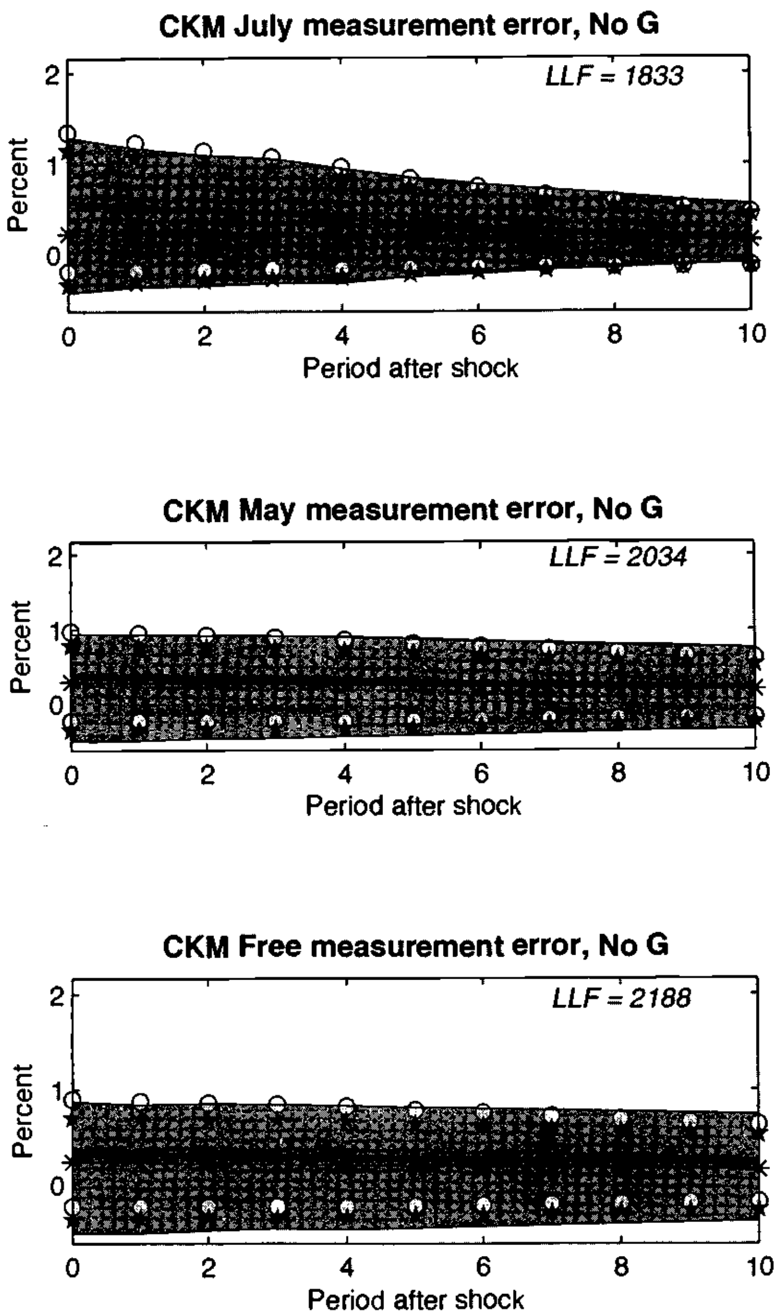


Figure 1.11 (continued)
The Treatment of CKM Measurement Error

specification of measurement error, the bias result remains even after relaxing CKM's assumption (compare the 1,1 and 1,2 graphs of figure 1.11). As noted above, the May specification of CKM's model has a likelihood that is orders of magnitude higher than the July specification. So, in the version of the CKM model selected by the likelihood criterion (i.e., the May version), the $\Delta \log G_t = \log z_t$ assumption plays a central role in driving the CKM's bias result.

In sum, CKM's examples, which imply that VARs with long-run identification display substantial bias, are not empirically interesting from a likelihood point of view. The bias in their examples is due to the way CKM choose the measurement error variance. When their measurement error specification is tested, it is overwhelmingly rejected in favor of an alternative in which the CKM bias result disappears.

Stochastic Process Uncertainty

CKM argue that there is considerable uncertainty in the business cycle literature about the values of parameters governing stochastic processes such as preferences and technology. They argue that this uncertainty translates into a wide class of examples in which the bias in structural VARs leads to severely misleading inference. The right panel in figure 1.12 summarizes their argument. The horizontal axis covers the range of values of $(\sigma_l / \sigma_z)^2$ considered by CKM. For each value of $(\sigma_l / \sigma_z)^2$ we estimate, by maximum likelihood, four parameters of the two-shock model: μ_z, τ_l, σ_l, and ρ_l.[29] We use the estimated model as a data generating process. The left vertical axis displays the small sample mean of the corresponding VAR-based estimator of the contemporaneous response of hours worked to a one-standard deviation technology shock.

Based on a review the RBC literature, CKM report that they have a roughly uniform prior over the different values of $(\sigma_l/\sigma_z)^2$ considered in figure 1.12. The figure indicates that for many of these values, the bias is large (compare the small sample mean, the solid line, with the true response, the starred line). For example, there is a noticeable bias in the two-shock CKM specification, where $(\sigma_l/\sigma_z)^2 = 1.1$.

We emphasize three points. First, as we stress repeatedly, bias cannot be viewed in isolation from sampling uncertainty. The two dashed lines in the figure indicate the 95 percent probability interval. These intervals are enormous relative to the bias. Second, not all values of $(\sigma_l / \sigma_z)^2$ are equally likely, and for the ones with greatest likelihood there is little bias. On the horizontal axis of the left panel of figure 1.12, we

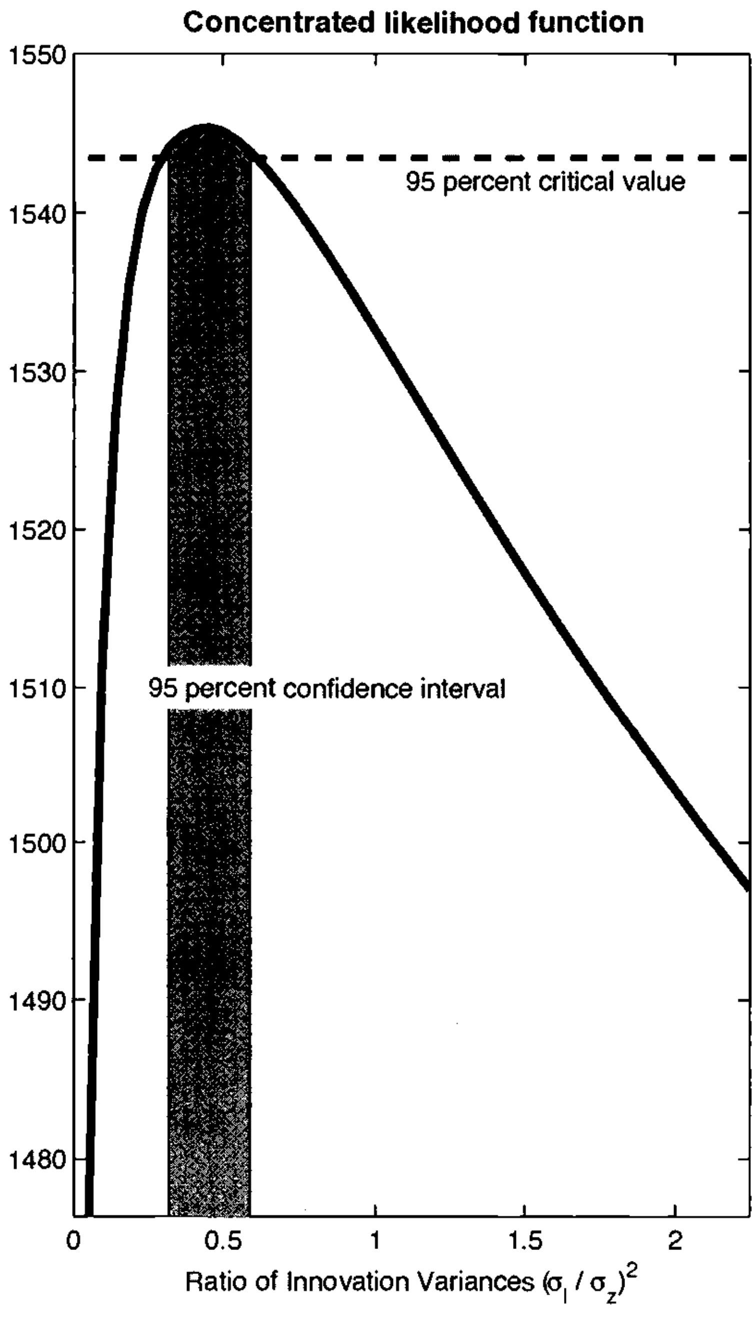


Figure 1.12
Stochastic Process Uncertainty

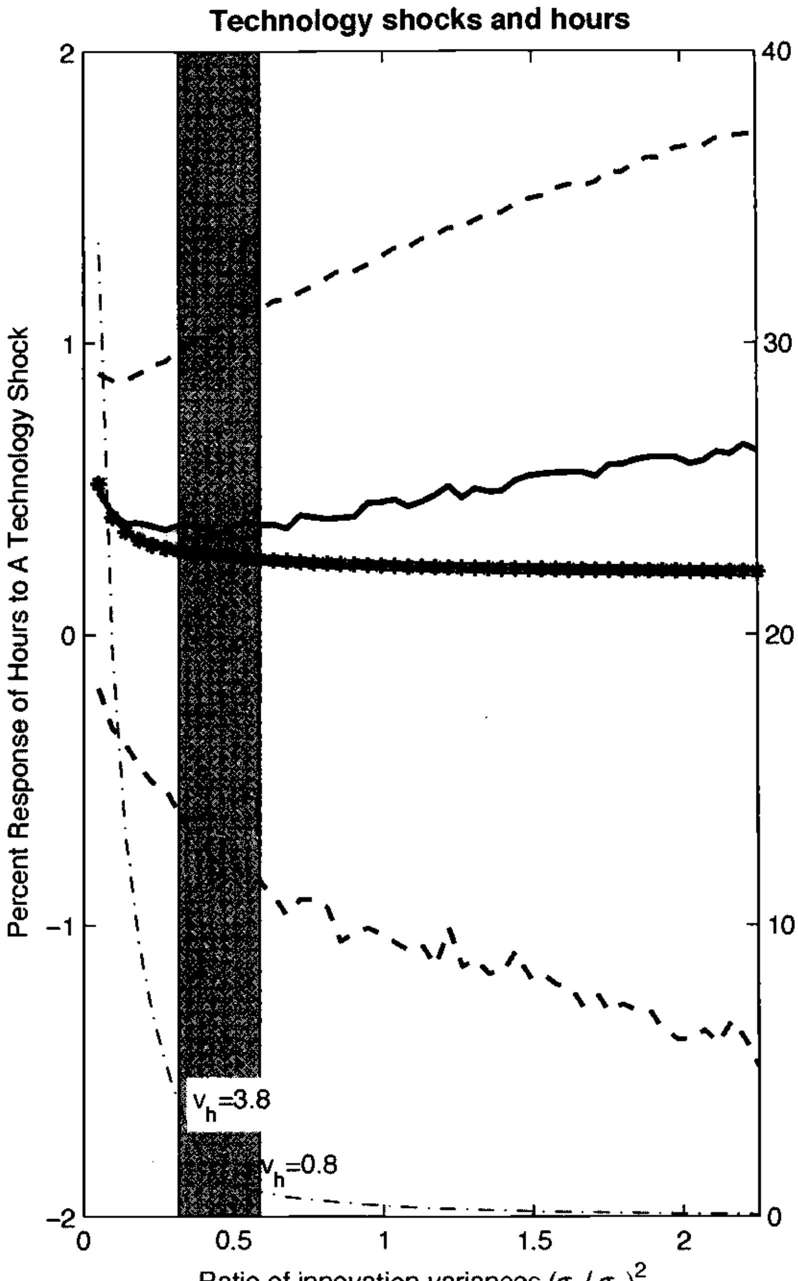


Figure 1.12 (continued)
Stochastic Process Uncertainty

display the same range of values of $(\sigma_l/\sigma_z)^2$ as in the right panel. On the vertical axis we report the log-likelihood value of the associated model. The peak of this likelihood occurs close to the estimated value in the two-shock MLE specification. Note how the log-likelihood value drops sharply as we consider values of $(\sigma_l/\sigma_z)^2$ away from the unconstrained maximum likelihood estimate. The vertical bars in the figure indicate the 95 percent confidence interval for $(\sigma_l/\sigma_z)^2$.[30] Figure 1.12 reveals that the confidence interval is very narrow relative to the range of values considered by CKM, and that within the interval, the bias is quite small.

Third, the right axis in the right panel of figure 1.12 plots V_h, the percent of the variance in log hours due to technology, as a function of $(\sigma_l/\sigma_z)^2$. The values of $(\sigma_l/\sigma_z)^2$ for which there is a noticeable bias correspond to model economies where V_h is less than 2 percent. Here, identifying the effects of a technology shock on hours worked is tantamount to looking for a needle in a haystack.

The Metric for Assessing the Performance of Structural VARs

CKM emphasize comparisons between the true dynamic response function in the data generating process and the response function that an econometrician would estimate using a four-lag VAR with an infinite amount of data. In our own analysis in section 4, we find population calculations with four lag VARs useful for some purposes. However, we do not view the probability limit of a four lag VAR as an interesting metric for measuring the usefulness of structural VARs. In practice econometricians do not have an infinite amount of data. Even if they did, they would certainly not use a fixed lag length. Econometricians determine lag length endogenously and, in a large sample, lag length would grow. If lag lengths grow at the appropriate rate with sample size, VAR-based estimators of impulse response functions are consistent. The interesting issue (to us) is how VAR-based procedures perform in samples of the size that practitioners have at their disposal. This is why we focus on small sample properties like bias and sampling uncertainty.

Over-Differencing

The potential power of the CKM argument lies in showing that VAR-based procedures are misleading, even under circumstances when everyone would agree that VARs should work well, namely when the econometrician commits no avoidable specification error. The econometrician does, however, commit one unavoidable specification error. The true VAR is infinite ordered, but the econometrician assumes the VAR has a finite number of lags. CKM argue that this seemingly innocuous assumption is fatal for VAR analysis. We have argued that this conclusion is unwarranted.

CKM present other examples in which the econometrician commits an avoidable specification error. Specifically, they study the consequences of over differencing hours worked. That is, the econometrician first differences hours worked when hours worked are stationary.[31] This error gives rise to bias in VAR-based impulse response functions that is large relative to sampling uncertainty. CKM argue that this bias is another reason not to use VARs.

However, the observation that avoidable specification error is possible in VAR analysis is not a problem for VARs per se. The possibility of specification error is a potential pitfall for any type of empirical work. In any case, CKM's analysis of the consequences of over differencing is not new. For example, Christiano, Eichenbaum, and Vigfusson (2003, hereafter, CEV) study a situation in which the true data generating process satisfies two properties: Hours worked are stationary and they rise after a positive technology shock. CEV then consider an econometrician who does VAR-based long-run identification when Y_t in (16) contains the growth rate of hours rather than the log level of hours. CEV show that the econometrician would falsely conclude that hours worked fall after a positive technology shock. CEV do not conclude from this exercise that structural VARs are not useful. Rather, they develop a statistical procedure to help decide whether hours worked should be first differenced or not.

CKM Ignore Short-Run Identification Schemes

We argue that VAR-based short-run identification schemes lead to remarkably accurate and precise inference. This result is of interest because the preponderance of the empirical literature on structural VARs explores the implications of short-run identification schemes. CKM are silent on this literature. McGrattan (2006) dismisses short-run identification schemes as "hokey." One possible interpretation of this adjective is that McGrattan can easily imagine models in which the identification scheme is incorrect. The problem with this interpretation is that all models are a collection of strong identifying assumptions, all of which can be characterized as "hokey." A second interpretation is that in McGrattan's view, the type of zero restrictions typically used in short run identification are not compatible with dynamic equilibrium theory. This view is simply incorrect (see Sims and Zha 2006). A third possible interpretation is that no one finds short-run identifying assumptions interesting. However, the results of short-run identification schemes have had an enormous effect on the construction of dynamic, general equilibrium models. See Woodford (2003) for a summary in the context of monetary models.

Sensitivity of Some VAR Results to Data Choices

CKM argue that VARs are very sensitive to the choice of data. Specifically, they review the papers by Francis and Ramey (2005), CEV, and Galí and Rabanal (2005), which use long-run VAR methods to estimate the response of hours worked to a positive technology shock. CKM note that these studies use different measures of per capita hours worked and output in the VAR analysis. The bottom panel of figure 1.13 displays the different measures of per capita hours worked that these studies use. Note how the low frequency properties of these series differ. The corresponding estimated impulse response functions and confidence intervals are reported in the top panel. CKM view it as a defect in VAR methodology that the different measures of hours worked lead to different estimated impulse response functions. We disagree. Empirical results *should* be sensitive to substantial changes in the data. A constructive response to the sensitivity in figure 1.13 is to carefully analyze the different measures of hours worked, see which is more appropriate, and perhaps construct a better measure. It is not constructive to dismiss an econometric technique that signals the need for better measurement.

CKM note that the principle differences in the hours data occur in the early part of the sample. According to CKM, when they drop these early observations they obtain different impulse response functions. However, as figure 1.13 shows, these impulse response functions are not significantly different from each other.

6 A Model with Nominal Rigidities

In this section we use the model in ACEL to assess the accuracy of structural VARs for estimating the dynamic response of hours worked to

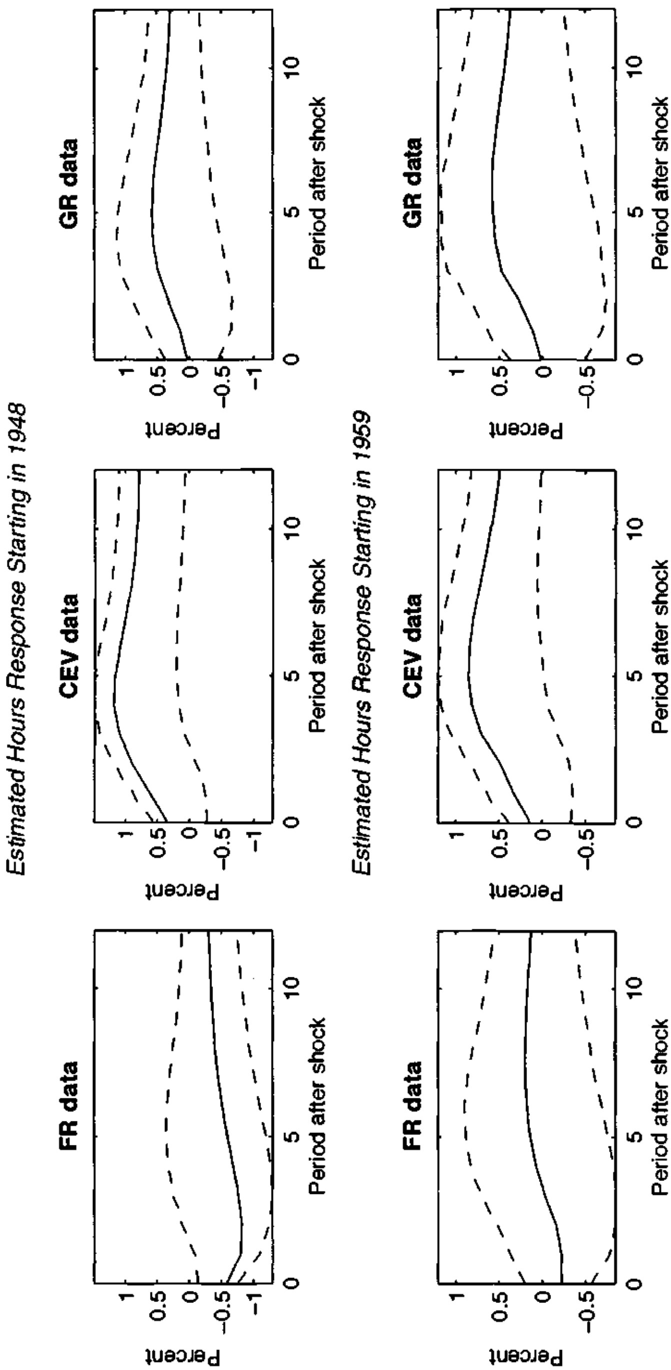


Figure 1.13
Data Sensitivity and Inference in VARs

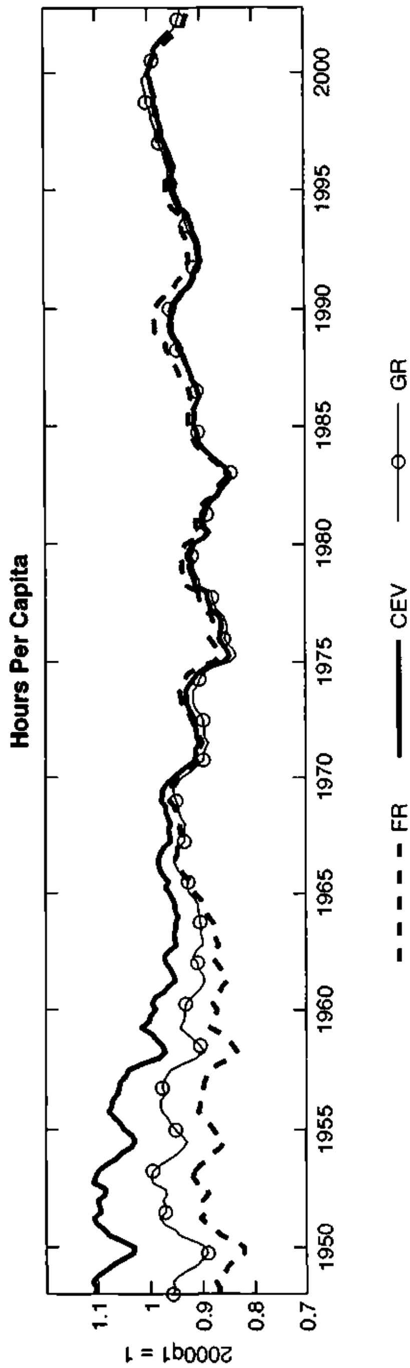


Figure 1.13 (continued)
Data Sensitivity and Inference in VARs

shocks. This model allows for nominal rigidities in prices and wages and has three shocks: a monetary policy shock, a neutral technology shock, and a capital-embodied technology shock. Both technology shocks affect labor productivity in the long run. However, the only shock in the model that affects the price of investment in the long run is the capital-embodied technology shock. We use the ACEL model to evaluate the ability of a VAR to uncover the response of hours worked to both types of technology shock and to the monetary policy shock. Our strategy for identifying the two technology shocks is similar to the one proposed by Fisher (2006). The model rationalizes a version of the short-run, recursive identification strategy used by Christiano, Eichenbaum, and Evans (1999) to identify monetary shocks. This strategy corresponds closely to the recursive procedure studied in section 2.3.2.

6.1 The Model

The details of the ACEL model, as well as the parameter estimates, are reported in Appendix A of the NBER Working Paper version of this paper. Here, we limit our discussion to what is necessary to clarify the nature of the shocks in the ACEL model. Final goods, Y_t, are produced using a standard Dixit-Stiglitz aggregator of intermediate goods, $y_t(i)$, $i \in (0, 1)$. To produce a unit of consumption goods, C_t, one unit of final goods is required. To produce one unit of investment goods, I_t, Υ_t^{-1} units of final goods are required. In equilibrium, Υ_t^{-1} is the price, in units of consumption goods, of an investment good. Let $\mu_{\Upsilon,t}$ denote the growth rate of Υ_t, let μ_Υ denote the nonstochastic steady state value of $\mu_{\Upsilon,t}$, and let $\hat{\mu}_{\Upsilon,t}$ denote the percent deviation of $\mu_{\Upsilon,t}$ from its steady state value:

$$\mu_{\Upsilon,t} = \frac{\Upsilon_t}{\Upsilon_{t-1}}, \quad \hat{\mu}_{\Upsilon,t} = \frac{\mu_{\Upsilon,t} - \mu_\Upsilon}{\mu_\Upsilon}. \tag{31}$$

The stochastic process for the growth rate of Υ_t is:

$$\hat{\mu}_{\Upsilon,t} = \rho_{\mu_\Upsilon} \hat{\mu}_{\Upsilon,t-1} + \sigma_{\mu_\Upsilon} \varepsilon_{\mu_\Upsilon,t}, \quad \sigma_{\mu_\Upsilon} > 0. \tag{32}$$

We refer to the i.i.d. unit variance random variable, $\varepsilon_{\mu_\Upsilon,t}$, as the capital-embodied technology shock. ACEL assume that the intermediate good, $y_t(i)$, for $i \in (0, 1)$ is produced using a Cobb-Douglas production function of capital and hours worked. This production function is perturbed by a multiplicative, aggregate technology shock denoted by Z_t. Let z_t denote the growth rate of Z_t, let z denote the nonstochastic steady state value of z_t, and let $\hat{z}_t$ denote the percentage deviation of z_t from its steady state value:

$$z_t = \frac{Z_t}{Z_{t-1}}, \quad \hat{z}_t = \frac{z_t - z}{z}. \tag{33}$$

The stochastic process for the growth rate of Z_t is:

$$\hat{z}_t = \rho_z \hat{z}_{t-1} + \sigma_z \varepsilon_t^z, \quad \sigma_z > 0, \tag{34}$$

where the i.i.d. unit variance random variable, ε_t^z, is the neutral shock to technology.

We now turn to the monetary policy shock. Let x_t denote M_t/M_{t-1}, where M_t denotes the monetary base. Let $\hat{x}_t$ denote the percentage deviation of x_t from its steady state, i.e., $(\hat{x}_t - x)/x$. We suppose that $\hat{x}_t$ is the sum of three components. One, $\hat{x}_{Mt}$, represents the component of $\hat{x}_t$ reflecting an exogenous shock to monetary policy. The other two, $\hat{x}_{zt}$ and $\hat{x}_{\Upsilon t}$, represent the endogenous response of $\hat{x}_t$ to the neutral and capital-embodied technology shocks, respectively. Thus monetary policy is given by:

$$\hat{x}_t = \hat{x}_{zt} + \hat{x}_{\Upsilon t} + \hat{x}_{Mt}. \tag{35}$$

ACEL assume that

$$\hat{x}_{M,t} = \rho_{xM} \hat{x}_{M,t-1} + \sigma_M \varepsilon_{M,t}, \quad \sigma_M > 0 \tag{36}$$

$$\hat{x}_{z,t} = \rho_{xz} \hat{x}_{z,t-1} + c_z \varepsilon_t^z + c_z^p \varepsilon_{t-1}^z$$

$$\hat{x}_{\Upsilon,t} = \rho_{x\Upsilon} \hat{x}_{\Upsilon,t-1} + c_\Upsilon \varepsilon_{\mu_{\Upsilon,t}} + c_\Upsilon^p \varepsilon_{\mu_{\Upsilon,t}}.$$

Here, $\varepsilon_{M,t}$ represents the shock to monetary policy and is an i.i.d. unit variance random variable.

Table 1.3 summarizes the importance of different shocks for the variance of hours worked and output. Neutral and capital-embodied technology shocks account for roughly equal percentages of the variance of hours worked (40 percent each), while monetary policy shocks account for the remainder. Working with HP-filtered data reduces the importance of neutral technology shocks to about 18 percent. Monetary policy shocks become much more important for the variance of hours worked. A qualitatively similar picture emerges when we consider output.

Table 1.3
Percent Contribution of Shocks in the ACEL model to the Variation in Hours and in Output

	Types of Shock		
Statistic	Monetary Policy	Neutral Technology	Capital-Embodied
Variance of logged hours	22.2	40.0	38.5
Variance of HP filtered logged hours	37.8	17.7	44.5
Variance of Δ_y	29.9	46.7	23.6
Variance of HP filtered logged output	31.9	32.3	36.1

Note: Results are average values based on 500 simulations of 3,100 observations each. ACEL: Altig Christiano, Eichenbaum, and Linde (2005).

It is worth emphasizing that neutral technology shocks are much more important in hours worked in the ACEL model than in the RBC model. This fact plays an important role in determining the precision of VAR-based inference using long-run restrictions in the ACEL model.

6.2 *Results*

We use the ACEL model to simulate 1,000 data sets each with 180 observations. We report results from two different VARs. In the first VAR, we simultaneously estimate the dynamic effect on hours worked of a neutral technology shock and a capital-embodied technology shock. The variables in this VAR are:

$$Y_t = \begin{pmatrix} \Delta \ln p_{It} \\ \Delta \ln a_t \\ \ln l_t \end{pmatrix},$$

where p_{It} denotes the price of capital in consumption units. The variable, $\ln (p_{It})$, corresponds to $\ln (\Upsilon_t^{-1})$ in the model. As in Fisher (2006), we identify the dynamic effects on Y_t of the two technology shocks, using a generalization of the strategy in section 2.3.1.[32] The details are provided in Appendix B of the NBER Working Paper version of this paper.

The 1,1 panel of figure 1.14 displays our results using the standard VAR procedure to estimate the dynamic response of hours worked to a neutral technology shock. Several results are worth emphasizing. First, the estimator is essentially unbiased. Second, the econometrician's estimator of sampling uncertainty is also reasonably unbiased. The

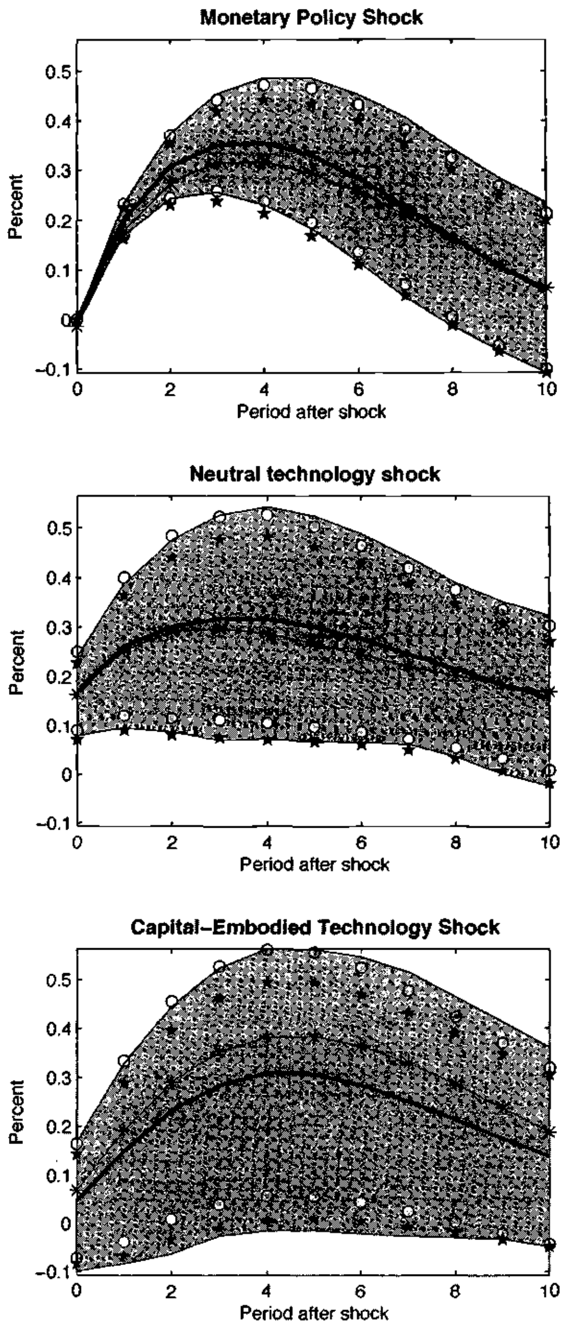


Figure 1.14
Impulse Response Results When the ACEL Model Is the DGP

circles and stars, which indicate the mean value of the econometrician's standard-deviation-based and percentile-based confidence intervals, roughly coincide with the boundaries of the gray area. However, there is a slight tendency, in both cases, to understate the degree of sampling uncertainty. Third, confidence intervals are small, relative to those in the RBC examples. Both sets of confidence intervals exclude zero at all lags shown. This result provides another example, in addition to the one provided by Erceg, Guerrieri, and Gust (2005), in which long-run identifying restrictions are useful for discriminating between models. An econometrician who estimates that hours drop after a positive technology shock would reject our parameterization of the ACEL model. Similarly, an econometrician with a model implying that hours fall after a positive technology shock would most likely reject that model if the actual data were generated by our parameterization of the ACEL model.

The 2,1 panel in figure 1.14 shows results for the response to a capital-embodied technology shock as estimated using the standard VAR estimator. The sampling uncertainty is somewhat higher for this estimator than for the neutral technology shock. In addition, there is a slight amount of bias. The econometrician understates somewhat the degree of sampling uncertainty.

We now consider the response of hours worked to a monetary policy shock. We estimate this response using a VAR with the following variables:

$$Y_t = \begin{pmatrix} \Delta \log a_t \\ \log l_t \\ R_t \end{pmatrix}.$$

As discussed in Christiano, Eichenbaum, and Evans (1999), the monetary policy shock is identified by choosing C to be the lower triangular decomposition of the variance covariance matrix, V, of the VAR disturbances. That is, we choose a lower triangular matrix, C with positive diagonal terms, such that $CC' = V$. Let $u_t = C\varepsilon_t$. We then interpret the last element of ε_t as the monetary policy shock. According to the results in the 1,2 panel of figure 1.14, the VAR-based estimator of the response of hours worked displays relatively little bias and is highly precise. In addition, the econometrician's estimator of sampling uncertainty is virtually unbiased. Suppose the impulse response in hours worked to a monetary policy shock were computed using VAR-based methods with data generated from this model. We conjecture that a model in which money is neutral, or in which a monetary expansion drives hours worked down, would be easy to reject.

7 Concluding Remarks

In this paper we study the ability of structural VARs to uncover the response of hours worked to a technology shock. We consider two classes of data generating processes. The first class consists of a series of real business cycle models that we estimate using maximum likelihood methods. The second class consists of the monetary model in ACEL. We find that with short-run restrictions, structural VARs perform remarkably well in all our examples. With long-run restrictions, structural VARs work well as long as technology shocks explain at least a very small portion of the variation in hours worked.

In a number of examples that we consider, VAR-based impulse response functions using long-run restrictions exhibit some bias. Even though these examples do not emerge from empirically plausible data generating processes, we find them of interest. They allow us to diagnose what can go wrong with long-run identification schemes. Our diagnosis leads us to propose a modification to the standard VAR-based procedure for estimating impulse response functions using long-run identification. This procedure works well in our examples.

Finally, we find that confidence intervals with long-run identification schemes are substantially larger than those with short-run identification schemes. In all empirically plausible cases, the VARs deliver confidence intervals that accurately reflect the true degree of sampling uncertainty. We view this characteristic as a great virtue of VAR-based methods. When the data contain little information, the VAR will indicate the lack of information. To reduce large confidence intervals the analyst must either impose additional identifying restrictions (i.e., use more theory) or obtain better data.

Acknowledgments

The first two authors are grateful to the National Science Foundation for Financial Support. We thank Lars Hansen and our colleagues at the Federal Reserve Bank of Chicago and the Board of Governors for useful comments at various stages of this project. The views in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or its staff.

Endnotes

1. See for example Sims (1989), Eichenbaum and Evans (1995), Rotemberg and Woodford (1997), Galí (1999), Francis and Ramey (2005), Christiano, Eichenbaum, and Evans (2005), and Del Negro, Schorfheide, Smets, and Wouters (2005).

2. See, for example, Basu, Fernald, and Kimball (2004), Christiano, Eichenbaum, and Vigfusson (2003, 2004), Fisher (2006), Francis and Ramey (2005), King, Plosser, Stock and Watson (1991), Shapiro and Watson (1988), and Vigfusson (2004). Francis, Owyang, and Roush (2005) pursue a related strategy to identify a technology shock as the shock that maximizes the forecast error variance share of labor productivity at a long but finite horizon.

3. This list is particularly long and includes at least Bernanke (1986), Bernanke and Blinder (1992), Bernanke and Mihov (1998), Blanchard and Perotti (2002), Blanchard and Watson (1986), Christiano and Eichenbaum (1992), Christiano, Eichenbaum, and Evans (2005), Cushman and Zha (1997), Eichenbaum and Evans (1995), Hamilton (1997), Rotemberg and Woodford (1992), Sims (1986), and Sims and Zha (2006).

4. See also Fernandez-Villaverdez, Rubio-Ramirez, and Sargent (2005) who investigate the circumstances in which the economic shocks are recoverable from the VAR disturbances. They provide a simple matrix algebra check to assess recoverability. They identify models in which the conditions are satisfied and other models in which they are not.

5. Let $\tilde{k}_t = k_t/Z_{t-1}$. Then, $\hat{k} = (\tilde{k}_t - \tilde{k})/\tilde{k}$, where $\tilde{k}$ denotes the value of $\tilde{k}_t$ in nonstochastic steady state.

6. For an early example, see Hansen and Sargent (1980, footnote 12). Sims and Zha (forthcoming) discuss the possibility that, although a given economic shock may not lie exactly in the space of current and past Y_t, it may nevertheless be "close." They discuss methods to detect this case.

7. Cooley and Dwyer (1998) argue that in the standard RBC model, if technology shocks have a unit root, then per capita hours worked will be difference stationary. This claim, which plays an important role in their analysis of VARs, is incorrect.

8. We implement the procedure for estimating C_2 by computing $CC' = V$, where C is the lower triangular Cholesky decomposition of V, and setting C_2 equal to the second column of C.

9. We use the standard Kalman filter strategy discussed in Hamilton (1994, section 13.4). We remove the sample mean from X_t prior to estimation and set the measurement error in the Kalman filter system to zero, i.e., $R = 0$ in (6).

10. See, for example, Christiano (1988), Christiano et al. (2004), and Smets and Wouters (2003).

11. We compute forecast error variances based on a four lag VAR. The variables in the VAR depend on whether the calculations correspond to the two or three shock model. In the case of the two-shock model, the VAR has two variables, output growth and log hours. In the case of the three-shock model, the VAR has three variables: output growth, log hours and the log of the investment to output ratio. Computing V_h requires estimating VARs in artificial data generated with all shocks, as well as in artificial data generated with only the technology shock. In the latter case, the one-step ahead forecast error from the VAR is well defined, even though the VAR coefficients themselves are not well defined due to multicollinearity problems.

12. When we measure V_h according to (1), V_h drops from 3.73 in the two-shock MLE model to 0.18 in the three-shock MLE model. The analogous drop in V_h is an order of magnitude smaller when V_h is measured using (2) or (3). The reason for this difference is that ρ_l goes from 0.986 in the two-shock MLE model to 0.9994 in the three-shock MLE model. In the latter specification there is a near-unit root in $\tau_{l,t}$, which translates into a near-unit root in hours worked. As a result, the variance of hours worked becomes very large at the low frequencies. The near-unit root in τ_{lt} has less of an effect on hours worked at high and business cycle frequencies.

13. Sims and Zha (1999) refer to what we call the percentile-based confidence interval as the "other-percentile bootstrap interval." This procedure has been used in several studies, such as Blanchard and Quah (1989), Christiano, Eichenbaum, and Evans (1999), Francis and Ramey (2005), McGrattan (2006), and Runkle (1987). The standard-deviation based confidence interval has been used by other researchers, such as Christiano, Eichenbaum, and Evans (2005), Galí (1999), and Galí and Rabanal (2005).

14. For each lag starting at the impact period, we ordered the 1,000 estimated impulse responses from smallest to largest. The lower and upper boundaries correspond to the 25^{th} and the 975^{th} impulses in this ordering.

15. An extreme example, in which the point estimates roughly coincide with one of the boundaries of the percentile-based confidence interval, appears in Blanchard and Quah (1989).

16. As σ_l falls, the total volatility of hours worked falls, as does the relative importance of labor tax shocks. In principle, both effects contribute to the decline in sampling uncertainty.

17. The minimization in (26) is actually over the trace of the indicated integral. One interpretation of (26) is that it provides the probability limit of our estimators—what they would converge to as the sample size increases to infinity. We do not adopt this interpretation, because in practice an econometrician would use a consistent lag-length selection method. The probability limit of our estimators corresponds to the true impulse response functions for all cases considered in this paper.

18. The derivation of this formula is straightforward. Write (10) in lag operator form as follows:

$$Y_t = B(L)Y_{t-1} + u_t,$$

where $Eu_t u_t' = V$. Let the fitted disturbances associated with a particular parameterization, $\hat{B}(L)$, be denoted $\hat{u}_t$. Simple substitution implies:

$$\hat{u}_t = [B(L) - \hat{B}(L)]Y_{t-1} + u_t.$$

The two random variables on the right of the equality are orthogonal, so that the variance $\hat{u}_t$ of is just the variance of the sum of the two:

$$var(\hat{u}_t) = var([B(L) - \hat{B}(L)]Y_{t-1}) + V.$$

Expression (26) in the text follows immediately.

19. By $V > \hat{V}$, we mean that $V - \hat{V}$ is a positive definite matrix.

20. In the earlier discussion it was convenient to adopt the normalization that the technology shock is the second element of ε_t. Here, we adopt the same normalization as for the long-run identification—namely, that the technology shock is the first element of ε_t.

21. This result explains why lag-length selection methods, such as the Akaike criterion, almost never suggest values of q greater than four in artificial data sets of length 180, regardless of which of our data generating methods we used. These lag length selection methods focus on $\hat{V}$.

22. Equation (26) shows that $\hat{B}(1)$ corresponds to only a single point in the integral. So other things equal, the estimation criterion assigns *no* weight at all to getting $\hat{B}(1)$ right. The reason $B(1)$ is identified in our setting is that the $B(\omega)$ functions we consider are continuous at $\omega = 0$.

23. A similar argument is presented in Ravenna (2005).

24. Christiano, Eichenbaum, and Vigfusson (2006) also consider the estimator proposed by Andrews and Monahan (1992).

25. The rule of always setting the bandwidth, r, equal to sample size does not yield a consistent estimator of the spectral density at frequency zero. We assume that as sample size is increased beyond $T = 180$, the bandwidth is increased sufficiently slowly to achieve consistency.

26. Because (26) is a quadratic function, we solve the optimization problem by solving the linear first-order conditions. These are the Yule-Walker equations, which rely on population second moments of the data. We obtain the population second moments by complex integration of the reduced form of the model used to generate the data, as suggested by Christiano (2002).

27. To ensure comparability of results we use CKM's computer code and data, available on Ellen McGrattan's webpage. The algorithm used by CKM to form the estimation criterion is essentially the same as the one we used to estimate our models. The only difference is that CKM use an approximation to the Gaussian function by working with the steady state Kalman gain. We form the exact Gaussian density function, in which the Kalman gain varies over dates, as described in Hamilton (1994). We believe this difference is inconsequential.

28. When generating the artificial data underlying the calculations in the 3,1 panel of figure 1.11, we set the measurement error to zero. (The same assumption was made for all the results reported here.) However, simulations that include the estimated measurement error produce results that are essentially the same.

29. We use CKM's computer code and data to ensure comparability of results.

30. The bounds of this interval are the upper and lower values of $(\sigma_l / \sigma_z)^2$ where twice the difference of the log-likelihood from its maximal value equals the critical value associated with the relevant likelihood ratio test.

31. For technical reasons, CKM actually consider "quasi differencing" hours worked using a differencing parameter close to unity. In small samples this type of quasi differencing is virtually indistinguishable from first differencing.

32. Our strategy differs somewhat from the one pursued in Fisher (2006), who applies a version of the instrumental variables strategy proposed by Shapiro and Watson (1988).

33. Similar specifications have been used by authors such as Sims (1994) and Schmitt-Grohé and Uribe (2004).

References

Altig, David, Lawrence Christiano, Martin Eichenbaum, and Jesper Linde. 2005. "Firm-Specific Capital, Nominal Rigidities, and the Business Cycle." NBER Working Paper Series 11034. Cambridge, MA: National Bureau of Economic Research.

Andrews, Donald W. K., and J. Christopher Monahan. 1992. "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica* 60: 953–966.

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are Technology Improvements Contractionary?" *American Economic Review* 96(5): 1418–1448.

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

Bernanke, Ben S., and Alan S. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82: 901–921.

Bernanke, Ben S., and Ilian Mihov. 1998. "Measuring Monetary Policy." *Quarterly Journal of Economics* 113: 869–902.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117: 1329–1368.

Blanchard, Olivier, and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* 79: 655–673.

Blanchard, Olivier, and Mark Watson. 1986. "Are Business Cycles All Alike?" In Robert Gordon, ed., *Continuity and Change in the American Business Cycle*. Chicago, IL: University of Chicago Press, 123–156.

Chari, V. V., Patrick J. Kehoe, and Ellen McGrattan. 2005a. "A Critique of Structural VARs Using Real Business Cycle Theory." Working Paper Series 631. Minneapolis, MN: Federal Reserve Bank of Minneapolis, May.

Chari, V. V., Patrick J. Kehoe, and Ellen McGrattan. 2005b. "A Critique of Structural VARs Using Real Business Cycle Theory." Working Paper Series 631. Minneapolis, MN: Federal Reserve Bank of Minneapolis, July.

Christiano, Lawrence J. 1988. "Why Does Inventory Investment Fluctuate So Much?" *Journal of Monetary Economics* 21: 247–280.

Christiano, Lawrence J. 2002. "Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients." *Computational Economics* 20: 21–55.

Christiano, Lawrence J., and Martin Eichenbaum. 1992. "Identification and the Liquidity Effect of a Monetary Policy Shock." In Alex Cukierman, Zvi Hercowitz, and Leonardo Leiderman, eds., *Political Economy, Growth, and Business Cycles*. Cambridge, MA: MIT Press, 335–370.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans. 1999. "Monetary Policy Shocks: What Have We Learned and to What End?" In John B.Taylor and Michael D. Woodford, eds., *Handbook of Macroeconomics*, Volume 1A, Amsterdam: Elsevier Science, 65–148.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113: 1–45.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson. 2003. "What Happens after a Technology Shock?" NBER Working Paper Series 9819. Cambridge, MA: National Bureau of Economic Research.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson. 2004. "The Response of Hours to a Technology Shock: Evidence Based on Direct Measures of Technology." *Journal of the European Economic Association* 2: 381–395.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson. 2006. "Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions." *Journal of the European Economic Association* 4(2-3): 475–483.

Cooley, Thomas F., and Mark Dwyer. 1998. "Business Cycle Analysis without Much Theory: A Look at Structural VARs." *Journal of Econometrics* 83: 57–88.

Cushman, David O., and Tao Zha. 1997. "Identifying Monetary Policy in a Small Open Economy under Flexible Exchange Rates." *Journal of Monetary Economics* 39: 433–448.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Raf Wouters. 2005. "On the Fit and Forecasting Performance of New Keynesian Models." Unpublished paper, University of Pennsylvania.

Eichenbaum, Martin, and Charles Evans. 1995. "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates." *Quarterly Journal of Economics* 110: 975–1009.

Erceg, Christopher J., Luca Guerrieri, and Christopher Gust. 2005. "Can Long-Run Restrictions Identify Technology Shocks?" *Journal of the European Economic Association* 3: 1237–1278.

Erceg, Christopher J., Dale W. Henderson, and Andrew T. Levin. 2000. "Optimal Monetary Policy with Staggered Wage and Price Contracts." *Journal of Monetary Economics* 46: 281–313.

Faust, Jon, and Eric Leeper. 1997. "When Do Long-Run Identifying Restrictions Give Reliable Results?" *Journal of Business and Economic Statistics* 15: 345–353.

Fernandez-Villaverde, Jesus, Juan F. Rubio-Ramirez, and Thomas J. Sargent. 2005. "A, B, C's (and D's) for Understanding VARs." Unpublished paper, New York University.

Fisher, Jonas. 2006. "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." *Journal of Political Economy* 114(3): 413–451.

Francis, Neville, Michael T. Owyang, and Jennifer E. Roush. 2005. "A Flexible Finite-Horizon Identification of Technology Shocks." Working Paper Series 2005–024A. St. Louis: Federal Reserve Bank of St. Louis.

Francis, Neville, and Valerie A. Ramey. 2005. "Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited." *Journal of Monetary Economics* 52: 1379–1399.

Galí, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89: 249–271.

Galí, Jordi, and Pau Rabanal. 2005. "Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data?" In Mark Gertler and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2004*. Cambridge, MA: MIT Press.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, James D. 1997. "Measuring the Liquidity Effect." *American Economic Review* 87: 80–97.

Hansen, Lars, and Thomas Sargent. 1980. "Formulating and Estimating Dynamic Linear Rational Expectations Models." *Journal of Economic Dynamics and Control* 2: 7–46.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. 1991. "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81: 819–840.

McGrattan, Ellen. 2006. "A Critique of Structural VARs Using Business Cycle Theory." Presentation at the annual meeting of the American Economic Association, Boston, January 6–8. Available at <http://minneapolisfed.org/research/economists/mcgrattan/CEV/assa.htm>.

Pagan, Adrian R., and John C. Robertson. 1998. "Structural Models of the Liquidity Effect." *Review of Economics and Statistics* 80: 202–217.

Ravenna, Federico. 2005. "Vector Autoregressions and Reduced Form Representations of Dynamic Stochastic General Equilibrium Models." Unpublished manuscript.

Rotemberg, Julio J., and Michael Woodford. 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy* 100: 1153–1207.

Rotemberg, Julio J., and Michael Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." In Ben S. Bernanke and Julio Rotemberg, eds., *NBER Macroeconomics Annual 1997*. Cambridge, MA: MIT Press.

Runkle, David E. 1987. "Vector Autoregressions and Reality." *Journal of Business and Economic Statistics* 5: 437–442.

Schmitt-Grohé, Stefanie, and Martin Uribe. 2004. "Optimal Fiscal and Monetary Policy under Sticky Prices." *Journal of Economic Theory* 114: 198–230.

Shapiro, Matthew, and Mark Watson. 1988. "Sources of Business Cycle Fluctuations." In Stanley Fischer, ed., *NBER Macroeconomics Annual 1988*. Cambridge, MA: MIT Press.

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

Sims, Christopher. 1980. "Macroeconomics and Reality." *Econometrica* 48: 1–48.

Sims, Christopher. 1986. "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis, *Quarterly Review* 10: 2–16.

Sims, Christopher. 1989. "Models and Their Uses." *American Journal of Agricultural Economics* 71: 489–494.

Sims, Christopher. 1994. "A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy." *Economic Theory* 4(3): 381–399.

Sims, Christopher, and Tao Zha. 1999. "Error Bands for Impulse Responses." *Econometrica* 67: 1113–1155.

Sims, Christopher, and Tao Zha. 2006. "Does Monetary Policy Generate Recessions?" *Macroeconomic Dynamics* 10(2): 231–272.

Smets, Frank, and Raf Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1: 1123–1175.

Vigfusson, Robert J. 2004. "The Delayed Response to a Technology Shock: A Flexible Price Explanation." International Finance Discussion Paper Series 2004–810. Washington, D.C.: Board of Governors of the Federal Reserve System.

Woodford, Michael M. 1996. "Control of the Public Debt: A Requirement for Price Stability?" NBER Working Paper Series no. 5684. Cambridge, MA.: National Bureau of Economic Research.

Woodford, Michael M. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ: Princeton University Press.

Appendix A A Model with Nominal Wage and Price Rigidities

This appendix describes the ACEL model used in section 6. The model economy is composed of households, firms, and a monetary authority.

There is a continuum of households, indexed by $j \in (0, 1)$. The j^{th} household is a monopoly supplier of a differentiated labor service, and sets its wage subject to Calvo-style wage frictions. In general, households earn different wage rates and work different amounts. A straightforward extension of arguments in Erceg, Henderson, and Levin (2000) and in Woodford (1996) establishes that in the presence of state contingent securities, households are homogeneous with respect to consumption and asset holdings. Our notation reflects this result. The preferences of the j^{th} household are given by:

$$E_t^j \sum_{l=0}^{\infty} \beta^{l-t} \left[\log(C_{t+1} - bC_{t+l-1}) - \psi_L \frac{h_{j,t+1}^2}{2} \right],$$

where $\psi_L \geq 0$ and E_t^j is the time t expectation operator, conditional on household j's time t information set. The variable, C_t, denotes time t consumption and h_{jt} denotes time t hours worked. The household's asset evolution equation is given by:

$$M_{t+1} = R_t\,[M_t - Q_t + (x_t - 1)M^a_t\,] + A_{j,t} + Q_t + W_{j,t}h_{j,t} + D_t - (1 + \eta\,(V_t))\,P_tC_t.$$

Here, M_t and Q_t denote, respectively, the household's stock of money, and cash balances at the beginning of period t. The variable $W_{j,t}$ represents the nominal wage rate at time t. In addition D_t and $A_{j,t}$ denote firm profits and the net cash inflow from participating in state-contingent security markets at time t. The variable, x_t, represents the gross growth rate of the economy-wide per capita stock of money, M^a_t. The quantity $(x_t - 1)M^a_t$ is a lump-sum payment to households by the monetary authority. The household deposits $M_t - Q_t + (x_t - 1)M^a_t$ with a financial intermediary. The variable, R_t, denotes the gross interest rate. The variable, V_t, denotes the time t velocity of the household's cash balances:

$$V_t = \frac{P_t C_t}{Q_t}, \tag{A1}$$

where $\eta(V_t)$ is increasing and convex.[33] For the quantitative analysis of our model, we must specify the level and the first two derivatives of the transactions function, $\eta(V)$, evaluated in steady state. We denote these by η, η', and η'', respectively. Let ε denote the interest semi-elasticity of money demand in steady state:

$$\varepsilon \equiv -\frac{100 \times d\log\left(\frac{Q_t}{P_t}\right)}{400 \times dR_t}.$$

Let V and η denote the values of velocity and $\eta(V_t)$ in steady state. ACEL parameterize the second-order Taylor series expansion of $\eta(\cdot)$ about steady state. The values of η, η', and η'', are determined by ACEL's estimates of ε, V, and η.

The j^{th} household is a monopoly supplier of a differentiated labor service, h_{jt}. It sells this service to a representative, competitive firm that transforms it into an aggregate labor input, L_t, using the technology:

$$H_t = \left[\int_0^1 h_{j,t}^{\frac{1}{\lambda_\omega}} dj\right]^{\lambda_\omega}, \quad 1 \le \lambda_\omega < \infty.$$

Let W_t denote the aggregate wage rate, i.e., the nominal price of H_t. The household takes H_t and W_t as given.

In each period, a household faces a constant probability, $1 - \xi_w$, of being able to re-optimize its nominal wage. The ability to re-optimize is independent across households and time. If a household cannot re-optimize its wage at time t, it sets W_{jt} according to:

$$W_{j,t} = \pi_{t-1}\mu_{z^*}W_{j,t-1},$$

where $\pi_{t-1} \equiv P_{t-1}/P_{t-2}$. The presence of μ_{z^*} implies that there are no distortions from wage dispersion along the steady state growth path.

At time t a final consumption good, Y_t, is produced by a perfectly competitive, representative final good firm. This firm produces the final good by combining a continuum of intermediate goods, indexed by $i \in [0, 1]$, using the technology

$$Y_t = \left[\int_0^1 y_t(i)^{\frac{1}{\lambda_f}} di\right]^{\lambda_f}, \tag{A2}$$

where $1 \leq \lambda_f < \infty$ and $y_t(i)$ denotes the time t input of intermediate good i. The firm takes its output price, P_t, and its input prices, $P_t(i)$, as given and beyond its control.

Intermediate good i is produced by a monopolist using the following technology:

$$y_t(i) = \begin{cases} K_t(i)^{\alpha}(Z_t h_t(i))^{1-\alpha} - \phi z_t^* & \text{if } K_t(i)^{\alpha}(Z_t h_t(i))^{1-\alpha} \geq \phi z_t^* \\ 0 & \text{otherwise} \end{cases} \tag{A3}$$

where $0 < \alpha < 1$. Here, $h_t(i)$ and $K_t(i)$ denote time t labor and capital services used to produce the i^{th} intermediate good. The variable Z_t represents a time t shock to the technology for producing intermediate output. The growth rate of Z_t, Z_t / Z_{t-1}, is denoted by μ_{zt}. The non-negative scalar, ϕ, parameterizes fixed costs of production. To express the model in terms of a stochastic steady state, we find it useful to define the variable z_t^* as:

$$z_t^* = \Upsilon_t^{\frac{\alpha}{1-\alpha}} Z_t, \tag{A4}$$

where Υ_t represents a time t shock to capital-embodied technology. The stochastic process generating Z_t is defined by (33) and (34). The stochastic process generating Υ_t is defined by (31) and (32).

Intermediate good firms hire labor in perfectly competitive factor markets at the wage rate, W_t. Profits are distributed to households at the end of each time period. We assume that the firm must borrow the wage bill in advance at the gross interest rate, R_t.

In each period, the i^{th} intermediate goods firm faces a constant probability, $1 - \xi_p$, of being able to re-optimize its nominal price. The ability to re-optimize prices is independent across firms and time. If firm i cannot re-optimize, it sets $P_t(i)$ according to:

$$P_t(i) = \pi_{t-1} P_{t-1}(i). \tag{A5}$$

Let $\bar{K}_t(i)$ denote the physical stock of capital available to the i^{th} firm at the beginning of period t. The services of capital, $K_t(i)$ are related to stock of physical capital, by:

$$\bar{K}_t(i) = u_t(i)\bar{K}_t(i).$$

Here $u_t(i)$ is firm i's capital utilization rate. The cost, in investment goods, of setting the utilization rate to $u_t(i)$ is $a(u_t(i))\,\bar{K}_t(i)$, where $a(\cdot)$ is increasing and convex. We assume that $u_t(i) = 1$ in steady state and $a(1) = 0$. These two conditions determine the level and slope of $a(\cdot)$ in steady state. To implement our log-linear solution method, we must also specify a value for the curvature of a in steady state, $\sigma_a = a''(1) / a'(1) \geq 0$.

There is no technology for transferring capital between firms. The only way a firm can change its stock of physical capital is by varying the rate of investment, $I_t(i)$, over time. The technology for accumulating physical capital by intermediate good firm i is given by:

$$F(I_t(i), I_{t-1}(i)) = \left(1 - S\left(\frac{I_t(i)}{I_{t-1}(i)}\right)\right) I_t(i),$$

where

$$\bar{K}_{t+1}(i) = (1-\delta)\bar{K}_t(i) + F(I_t(i), I_{t-1}(i)).$$

The adjustment cost function, S, satisfies $S = S' = 0$, and $S'' > 0$ in steady state. Given the log-linearization procedure used to solve the model, we need not specify any other features of the function S.

The present discounted value of the i^{th} intermediate good's net cash flow is given by:

$$E_t \sum_{j=0}^{\infty} \beta^j \upsilon_{t+j} \{ P_{t+j}(i) y_{t+j}(i) - R_{t+j} W_{t+j} h_t(i) - P_{t+j} \Upsilon_{t+j}^{-1} [I_{t+j}(i) \tag{A6}$$

$$+ a(u_{t+j}(i)) \bar{K}_{t+j}(i)]\},$$

where R_t denotes the gross nominal rate of interest.

The monetary policy rule is defined by (35) and (36). Financial intermediaries receive $M_t - Q_t + (x_t - 1)M_t$ from the household. Our notation reflects the equilibrium condition, $M_t^a = M_t$. Financial intermediaries lend all of their money to intermediate good firms, which use the funds to pay labor wages. Loan market clearing requires that:

$$W_t H_t = x_t M_t - Q_t. \tag{A7}$$

The aggregate resource constraint is:

$$(1 + \eta(V_t))C_t + \Upsilon_t^{-1}[I_t + a(u_t)\bar{K}_t] \leq Y_t. \tag{A8}$$

We refer the reader to ACEL for a description of how the model is solved and for the methodology used to estimate the model parameters. The data and programs, as well as an extensive technical appendix, may be found at the following website: www.faculty.econ.northwestern.edu/faculty/christiano/research/ACEL/acelweb.htm.

Appendix B Long-Run Identification of Two Technology Shocks

This appendix generalizes the strategy for long-run identification of one shock to two shocks, using the strategy of Fisher (2006). As before, the VAR is:

$$Y_{t+1} = B(L)\, Y_t + u_t,\ \ Eu_tu'_t = V,$$

$$B(L) \equiv B_1 + B_2L + \ldots + B_qL^{q-1},$$

We suppose that the fundamental shocks are related to the VAR disturbances as follows:

$$u_t = C\varepsilon_t,\ E\varepsilon_t\varepsilon_t' = I,\ CC' = V,$$

where the first two element in ε_t are $\varepsilon_{\mu\Upsilon,t}$ and ε_t^z, respectively. The exclusion restrictions are:

$$\lim_{j\to\infty}[\tilde{E}_t a_{t+j} - \tilde{E}_{t-1}a_{t+j}] = f_z(\varepsilon_{\mu_{\Upsilon,t}}, \varepsilon_t^z, \text{ only})$$

$$\lim_{j\to\infty}[\tilde{E}_t \log p_{I,t+j} - \tilde{E}_{t-1}\log p_{I,t+j}] = f_\Upsilon(\varepsilon_{\mu_{\Upsilon,t}}, \text{ only}).$$

That is, only technology shocks have a long-run effect on the log-level of labor productivity, whereas only capital-embodied shocks have a long-run effect on the log-level of the price of investment goods. According to the sign restrictions, the slope of f_z with respect to its second argument and the slope of f_Υ are non-negative. Applying a suitably modified version of the logic in section 2.3.1, we conclude that, according to the exclusion restrictions, the indicated pattern of zeros must appear in the following 3 by 3 matrix:

$$[I - B(1)]^{-1}C = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ \text{number} & \text{number} & \text{number} \end{bmatrix}$$

The sign restrictions are $a, c > 0$. To compute the dynamic response of Y_t to the two technology shocks, we require the first two columns of C. To obtain these, we proceed as follows. Let $D \equiv [I - B(1)]^{-1}\, C$, so that:

$$DD' = [I - B(1)]^{-1}\, V\, [I - B(1)']^{-1} = S_Y(0)\,, \tag{B1}$$

where, as before, $S_Y(0)$ is the spectral density of Y_t at frequency-zero, as implied by the estimated VAR. The exclusion restrictions require that D have the following structure:

$$D = \begin{bmatrix} d_{11} & 0 & 0 \\ d_{21} & d_{22} & 0 \\ d_{31} & d_{32} & d_{33} \end{bmatrix}.$$

Here, the zero restrictions reflect our exclusion restrictions, and the sign restrictions require $d_{11}, d_{22} \geq 0$. Then,

$$DD' = \begin{bmatrix} d_{11}^2 & d_{11}d_{21} & d_{11}d_{31} \\ d_{21}d_{11} & d_{21}^2 + d_{22}^2 & d_{21}d_{31} + d_{22}d_{32} \\ d_{31}d_{11} & d_{31}d_{21} + d_{32}d_{22} & d_{31}^2 + d_{32}^2 + d_{33}^2 \end{bmatrix} = \begin{bmatrix} S_Y^{11}(0) & S_Y^{21}(0) & S_Y^{31}(0) \\ S_Y^{21}(0) & S_Y^{22}(0) & S_Y^{32}(0) \\ S_Y^{31}(0) & S_Y^{32}(0) & S_Y^{33}(0) \end{bmatrix}$$

and

$$d_{11} = \sqrt{S_Y^{11}(0)}, \quad d_{21} = S_Y^{21}(0)/d_{11}, \quad d_{31} = S_Y^{31}(0)/d_{11}$$

$$d_{22} = \sqrt{\frac{S_Y^{11}(0)S_Y^{22}(0) - (S_Y^{21}(0))^2}{S_Y^{11}(0)}}, \quad d_{32} = \frac{S_Y^{32}(0) - S_Y^{21}(0)S_Y^{31}(0)/d_{11}^2}{d_{22}}.$$

The sign restrictions imply that the square roots should be positive. The fact that $S_Y(0)$ is positive definite ensures that the square roots are real numbers. Finally, the first two columns of C are calculated as follows:

$$[C_1 \vdots C_2] = [I - B(1)][D_1 \vdots D_2],$$

where C_i is the i^{th} column of C and D_i is the i^{th} column of D, $i = 1, 2$.

To construct our modified VAR procedure, simply replace $S_Y(0)$ in ($B1$) by (29).

Comment

***Patrick J. Kehoe**, Federal Reserve Bank of Minneapolis, University of Minnesota, and NBER*

1 Introduction

Most of the existing structural VAR (SVAR) literature argues that a useful way of advancing theory is to directly compare impulse responses from structural VARs run on the data to theoretical impulse responses from models. The crux of the Chari, Kehoe, and McGrattan (2006) (henceforth, CKM) critique of this *common approach* is that it compares the empirical impulse responses from the data to inappropriate objects in the model. We argue that logically, instead of being compared to the theoretical impulse responses, the empirical impulse responses should be compared to impulse responses from identical structural VARs run on data from the model of the same length as the actual data. We refer to this latter approach as the *Sims-Cogley-Nason approach* since it has been advocated by Sims (1989) and successfully applied by Cogley and Nason (1995).

CKM argue that in making the inappropriate comparison, the common approach makes an error avoided by the Sims-Cogley-Nason approach. That error makes the common approach prone to various pitfalls, including small-sample bias and lag-truncation bias. For example, the data length may be so short that the researcher is forced to use a short lag length, and the estimated VAR may be a poor approximation to the model's infinite-order VAR. The Sims-Cogley-Nason approach avoids such problems because it treats the data from the U.S. economy and the model economy symmetrically.

On purely logical grounds, then, the Sims-Cogley-Nason approach seems to dominate the common approach.[1] How well does the common approach do in practice using SVARs based on long-run restrictions on data from a real business cycle model? CKM show that for data of the relevant length, SVARs do miserably: The bias is large and SVARs are unable to distinguish between models of interest—unless technology shocks account for virtually all the fluctuations in output.

Christiano, Eichenbaum, and Vigfusson (2006) (henceforth, CEV), perhaps the most prominent defenders of the common approach, seem to agree with CKM on the most important matters of substance. Indeed, since there seems to be no dispute that the Sims-Cogley-Nason approach dominates the common approach, there should be little disagreement over how future research in this area should be conducted. Likewise, there seems to be no dispute that when shocks other than technology play a sizable role in output fluctuations, SVARs do miserably. The primary point of disagreement between CEV and CKM is thus a relatively minor one about the likely size of the errors in the past literature that uses the common approach. CEV argue that the errors are small because the evidence is overwhelming that in U.S. data, technology shocks account for virtually all the fluctuations in output. CKM point to both 20 years of business cycle research and simple statistics in the data that all lead to the opposite conclusion about technology shocks and, hence, to the opposite conclusion as to the size of the errors of the common approach.

CEV also venture beyond the confines of the CKM critique and analyze SVARs with short-run restrictions. They focus on SVARs applied to monetary models which satisfy the same recursive identifying assumptions as their SVARs. CEV argue that the error in this application of the common approach is small, and thus the technique can be used broadly to distinguish promising models from the rest. Here the primary problem with their analysis is that it is subject to the Lucas and Stokey critique (Lucas and Stokey 1987): Only a tiny subset of existing monetary models in the literature actually satisfies the recursive identifying assumptions. That subset does not include even, for example, the best-known monetary models of Lucas (1972, 1990). Yet the technique has been used to reject these and other such models. Clearly, comparing impulse responses from SVARs with a set of identifying assumptions to those from models which do not satisfy those assumptions is problematic.

Notice that the Sims-Cogley-Nason approach is immune to the Lucas and Stokey critique. Under this approach, it is entirely coherent to compare impulse responses with a set of identifying assumptions to those from models which do not satisfy these assumptions. Under this approach, the impulse responses are simply statistics with possibly little economic interpretation. Now, those statistics may not be interpretable as being close to the model's theoretical response, but so what? When Kydland and Prescott (1982) compare variances, covariances, and cross-correlations in the model and the data, it does not matter whether these statistics have some deep economic interpretation.

Of course, it is not true that all statistics are equally desirable. What properties lead certain statistics to be more desirable than others? One important property is that the statistics vary across alternative models in such a way that, with samples of the lengths we have, they can be used to point with confidence toward one class of models and away from another. (If no such statistics exist, then the data have little to say about the theories of interest.) A second desirable property is that the statistics depend on key features of theory and not on inessential auxiliary assumptions. An important question for a serious assessment of the SVAR literature is, in what sense are the SVAR statistics more or less desirable than a host of other non–SVAR-related statistics? Regrettably, little or no work in the SVAR literature seems directed at this critical question.

To reiterate: The CKM critique does not apply to all SVAR analyses, only those that use the common approach rather than the Sims-Cogley-Nason approach. For most analyses, switching to that dominant approach would cost little—changing only a few lines of computer code and a few lines of text. By making such a switch, researchers using structural VARs can vastly enhance the role of VARs in guiding theory.

In these comments, I begin by carefully describing the difference between the common approach and the Sims-Cogley-Nason approach. Then I describe four issues of perceived disagreement between CKM and CEV about SVARs with long-run restrictions. Finally, in terms of CEV's analysis with short-run restrictions, I describe two critiques which need to be addressed by researchers who steadfastly refuse to abandon the common approach.

2 Getting Precise

Let me begin with some notation with which I can make the CKM argument precise.

The first step in both SVAR approaches is to run a VAR with p lags on a data set $\{Y_t\}_{t=1}^{T}$ and then apply the identifying assumptions to construct the impulse response matrices $A_i(p, T)$ for $i = 0, 1, \ldots$, where i denotes periods after the impact period and the notation emphasizes that the impulse responses depend on the lag length p and the sample size T. In applications using postwar U.S. data, it is common to set $p = 4$ and to have $T = 180$ or numbers similar to these, and I will denote the resulting matrices by $A_i^{us}(p = 4, T = 180)$.

The common approach emphasizes the interpretation of these matrices. For instance, in the standard example, the data consist of a measure of labor productivity and a measure of hours and the theoretical model has two shocks, technology and non-technology shocks. The first column of the impact matrix $A_0^{us}(p = 4, T = 180)$ is interpreted as the impact effect of the technology shock on productivity and hours, while the second column is interpreted as the impact effect of the non-technology shock on productivity and hours. The subsequent matrices are similarly interpreted.

In contrast, CKM and Sims, Cogley, and Nason view these matrices as moments of the data that may be used in discriminating among models of interest.

Now suppose we have a quantitative economic model in which the impulse responses to the technology and non-technology shocks are the matrices $D_i(\theta)$, $i = 0, 1, \ldots$, where θ denotes the model parameters. The second step of the common approach compares

$$A_i^{us}(p = 4, T = 180) \text{ to } D_i(\theta). \tag{1}$$

Sometimes this comparison is informal and implicit, as in the work of Galí (1999), Francis and Ramey (2005), and Galí and Rabanal (2005), who find that labor falls after a positive productivity shock and conclude that real business cycles are dead. Sometimes this comparison is formal and explicit, as in the work of Altig et al. (2005), and is used to choose model parameters θ.

The second step of the Sims-Cogley-Nason approach is quite different. To understand this step, let $\bar{A}_i(p,T \mid \theta)$ denote the mean of impulse responses found by applying the SVAR approach with p lags in the VAR to the many simulations of data of length T generated from the model with parameters θ. The second step of the Sims-Cogley-Nason approach compares

$$A_i^{US}(p = 4, T = 180) \;\text{ to }\; \bar{A}_i(p = 4, T = 180 | \theta). \tag{2}$$

At a conceptual level, we interpret the Sims-Cogley-Nason approach as advocating comparing the exact small-sample distribution of the estimator of the impulse responses with $p = 4$ and $T = 180$ to the estimated impulse response parameters. We view the simulations involved as a simple way to approximate that small-sample distribution. If it were feasible to analytically work out the small-sample distribution of the estimator, then so much the better.

CKM interpret (2) as the correct comparison, which is firmly grounded in (simulated) method-of-moments theory, and (1) as simply a mistake of the common approach.

The whole point of the CKM work is to quantify when and why these two comparisons will yield different answers, that is, when and why the two objects computed from the model, $\bar{A}_i(p = 4, T = 180 \mid \theta)$ and $D_i(\theta)$, will differ. Part of CKM's analysis focuses on the two-variable case with $Y_t(\alpha) = (\Delta(y_t/l_t), l_t - \alpha l_{t-1})'$, where y_t is the log of output, l_t is the log of hours, and $\alpha \in [0, 1]$ is the quasi-differencing parameter. The specification $Y_t(\alpha)$ nests three cases of interest: $\alpha = 0$, the level SVAR (LSVAR) case; $\alpha = 1$, the differenced SVAR (DSVAR) case; and $\alpha = .99$, the quasi-differenced SVAR (QDSVAR) case.

When θ is such that technology shocks do not account for the vast bulk of fluctuations in output, LSVARs do miserably: The bias is large and the confidence bands are so enormous that the technique is unable to distinguish among most classes of models of interest.

With such a θ, the DSVARs and QDSVARs also fare poorly: The bias is large enough to flip the sign of the impact coefficient of hours on a technology shock. While the confidence bands are large, they don't stop a researcher from rejecting that the simulated data came from a real business cycle model, even though they did. CKM think that this result suggests that researchers who have determined that real business cycle models are dead based on SVAR evidence may have come to that conclusion simply because they were not comparing the appropriate objects in the model and the data.

Note that, at least for the long-run restriction branch of the SVAR litterature, the issue is all about approximation error. If we had an infinite sample of data from a model that satisfies the identifying restrictions and we estimated a VAR with an infinite number of lags, we would have (in the relevant sense of convergence)

$$\bar{A}_i(p = \infty, T = \infty) = D_i(\theta) \tag{3}$$

for both the LSVAR and the QDSVAR cases, where, for simplicity, we have assumed that the identifying assumptions are sufficient as well as necessary. (As we discuss below, Marcet (2005) shows why (3) holds even for the DSVAR case in which hours are "over-differenced.")

With (1)–(3) in mind, note that $\bar{A}_i(p = 4, T = 4) - D_i(\theta)$ can be interpreted as the error in the common approach relative to the Sims-Cogley-Nason approach. CKM decompose this error into

$$[\bar{A}_i(p=4,T=180)-\bar{A}_i(p=4,T=\infty)]+[\bar{A}_i(p=4,T=\infty)-D_i(\theta)],$$

where the first term is the Hurwicz-type *small-sample bias* and the second term is the *lag-truncation bias*. It turns out that for both the LSVAR case and the QDSVAR case, most of the error is coming from the lag-truncation bias. Intuitively, this truncation bias arises both because the $p = 4$ specification forced terms to be zero that are not and because the OLS estimator adjusts the estimates of the included lags to compensate for those that have been excluded. CKM develop propositions that give intuition for when the error from the lag-truncation bias will be large.[2]

3 The Common Approach with Long-Run Restrictions

The SVAR literature with long-run restrictions, in general, and CEV, in particular, claim that the common approach is a state-of-the-art technique which is a useful guide for theory. We disagree. Here I describe three specific points of disagreement relevant to CEV's discussion of long-run restrictions and one point in which CEV seem to think there is disagreement where none really exists. My overall point here is that we all agree there exist circumstances under which the errors from using the common approach are small; however, as CKM have shown, these circumstances are not general. Moreover, regardless of the circumstances, this approach is dominated by what we consider the state-of-the-art technique, the Sims-Cogley-Nason approach. This approach is at least as easy to use as the common approach, and it has the advantage of a firm logical and statistical foundation.

Consider now the four points.

First, CEV argue that LSVARs are useful in guiding theory about fluctuations in the U.S. economy because in U.S. data, they say, technology shocks account for almost all of the fluctuations in output. We argue that while some reasonable statistics do point to technology shocks playing an overwhelming role, a number of other sensible statistics, as well as much of the literature, strongly suggest that their role is modest.

Second, CEV argue that even if technology shocks do not account for almost all of the fluctuations in output, there is a new estimator of impulse responses that virtually eliminates the bias associated with the standard OLS estimator. We argue that while for some parameter values this new estimator improves on the OLS estimator, for others it does worse. In this sense, the new estimator does not solve all the problems facing this literature.

Third, CEV ignore the DSVAR literature on the grounds, they say, that the DSVAR is misspecified because it incorrectly differences hours. This misspecification, they say, leads to incorrect estimates of impulse responses even with an infinite amount of data. We argue that here, for all practical purposes, CEV are wrong about the DSVAR being misspecified. Instead the only error in the DSVAR literature is the same as in the LSVAR literature: Using the common approach rather than the Sims-Cogley-Nason approach.

Finally, I consider a point on which there is actually no disagreement. CEV argue that when more variables are added to an LSVAR, in special cases it can sometimes usefully distinguish between classes of models. CEV somehow seem to think we disagree here, but we do not. Indeed, part of the point of CKM is to provide a theorem as to when LSVARs can and cannot perform this function. We emphasize, however, that the "can" circumstances are somewhat narrow.

3.1 Do Technology Shocks Account for Virtually All of the Fluctuations in Output?

CKM show that if technology shocks account for virtually all of the fluctuations in output, then the errors associated with the common approach are relatively small. Much of CEV's work is devoted to arguing that the U.S. data definitively show that technology shocks account for the vast bulk of the movements in output and non-technology shocks, almost none. There is a vast literature on this subject, much of it contradicting that stand.

Let's take a closer look at the issues at stake. Using the notation of CKM and ignoring means, we can write the stochastic processes for a technology shock, $\log Z_t$, and a non-technology shock, τ_{lt}, for both CEV and CKM, as

$$\log Z_{t+1} = \log Z_t + \log z_{t+1} \tag{4}$$

$$\tau_{lt+1} = \rho\tau_{lt} + \varepsilon_{lt+1}, \tag{5}$$

where log z_t and ε_{lt} are independent, mean zero, i.i.d. normal random variables with variances σ_z^2 and σ_l^2 and ρ is the serial correlation of the non-technology shock. Note that these stochastic processes are determined by three parameters (σ_z^2, σ_l^2, ρ). CEV estimate these parameters to be $\sigma_z^2 = (.00953)^2$, $\sigma_l^2 = (.0056)^2$, and $\rho = .986$. CKM show that the impulse errors in the SVARs increase with the ratio of the variances of the innovations σ_l^2/σ_z^2.

CEV's finding that LSVARs do well with U.S. data rests crucially on their estimate of the variance of non-technology shocks. CKM and CEV agree that LSVARs do miserably when this variance is large. The main disagreement between us here is whether we can confidently assert that, when the U.S. data are viewed through the lens of a real business cycle model, the variance of non-technology shocks is, indeed, small. CEV do not make clear that at a mechanical level, the only source of their disagreement with us is the relevant values of that one parameter σ_l^2. Here, to demonstrate that point, I set all of the parameters, except σ_l^2, equal to those of CEV.

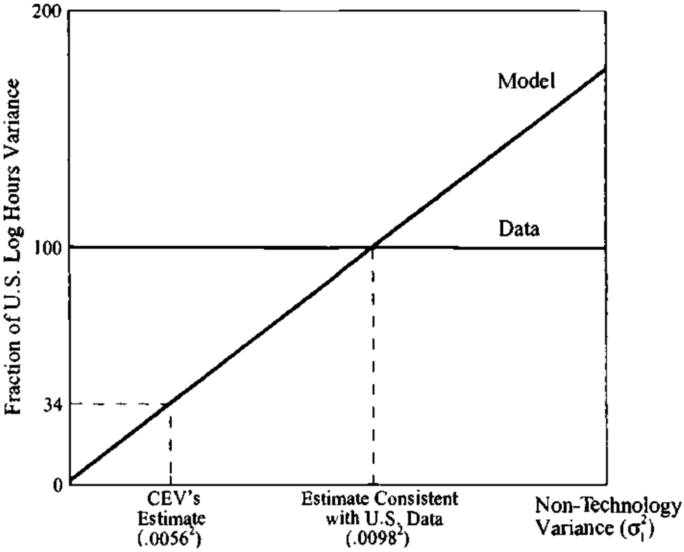
The question then is, what is a reasonable value for the variance of non-technology shocks? Before confronting this question formally, recall a well-known fact: In real business cycle models with unit root technology shocks, the volatility of hours due to technology shocks is tiny. The reason is that the unit root nature of the shocks diminishes the already small intertemporal substitution effects present in real business cycle models with mean-reverting shocks.[3] Indeed, based on unfiltered series in both the data and the model along with the CEV estimates for σ_z^2, we find that

$$\frac{\text{the variance of hours in the model with only technology shocks}}{\text{the variance of hours in the U.S. data}} = 1.8\%, \tag{6}$$

where for the hours series we use the same Prescott and Ueberfeldt series as in CKM. Thus, for the CEV model to reproduce the observed volatility of hours, the non-technology shocks alone must account for over 98 percent of the volatility in hours. In this sense, the data clearly suggest that non-technology shocks must be very large relative to technology shocks.

How large? To answer that question, in the top graph of figure 1.15, I plot the variance of hours in the model and in the data against the variance of the non-technology shocks, holding fixed σ_z^2 and ρ_l at CEV's values. Clearly, under these conditions, as σ_l^2 is increased, the variance of hours in the model rises. This graph shows that at CEV's estimate

Fraction of U.S. Hours Variance Generated by Model vs. Variance of Non-Technology Shocks



Impact Error and 95% Bootstrapped Confidence Bands

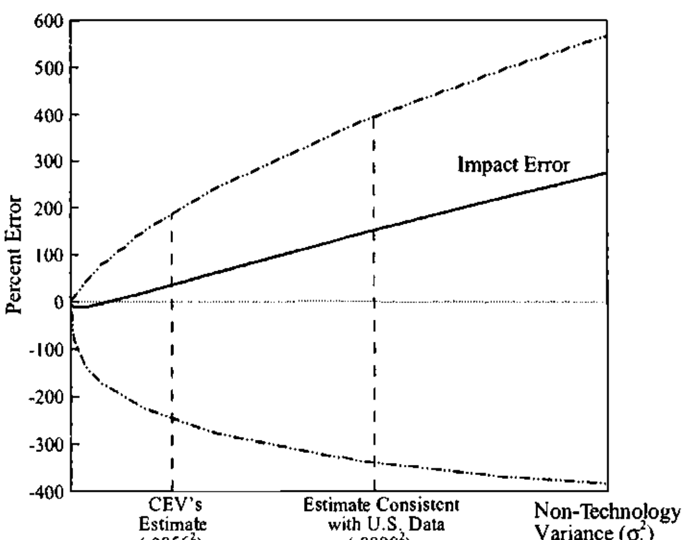


Figure 1.15
The Data Say Non-Technology Shocks Must Be Large, and When They Are, So Are the Bias and Confidence Bands

for σ_l^2 (.0056^2), hours are only about a third as volatile in their model as in the data. The graph also shows that for the model to account for the observed variability in hours, σ_l^2 must be substantially larger (about .0098^2).[4]

The bottom graph of figure 1.15 shows that when the parameter under dispute, σ_l^2, is chosen to reproduce CEV's estimate, the bias is modest but the confidence bands are large. When this parameter is chosen to reproduce the observed volatility of hours, the LSVAR does miserably: The bias is large and the confidence bands are so enormous that the technique is unable to distinguish among most classes of models of interest.

I should be clear that we do not disagree that there exist some statistics, including some maximum likelihood statistics, that would lead to the conclusion that non-technology shocks are small. CKM find that the maximum likelihood estimates are sensitive to the variables included in the observer equation, especially investment. Under some specifications, the variance of non-technology shocks is large while in others it is small. The reason for this sensitivity is that a stripped-down model like ours cannot mimic well all of the comovements in U.S. data, so that it matters what features of the data the researcher wants to mimic. In such a circumstance, we think it makes sense to use a limited-information technique in which we can choose the moments we want the model to do well on.

Therefore, in designing a laboratory to test whether the SVAR methodology works, we asked, what would be some desirable features of the data for the model to reproduce? We came up with three answers, all of which contradict the condition necessary for SVARs to work well in practice; that is, all three suggest that non-technology shocks must be large.

One of our answers, which motivates the exercise just conducted, is that if the whole point of the procedure is to decompose the movements in hours, then the model should generate volatility in hours similar to that in the data. As CKM demonstrate, in the context of the CEV model, to do that the model needs large non-technology shocks.

A second answer is that the laboratory model should reproduce the key statistic that both started off the whole debate and is the main result in the long-run restriction SVAR literature: Galí's (1999) initial drop in hours after a positive technology shock. CKM ask, holding fixed the estimates of the variance of technology shocks and the persistence of non-technology shocks, what must be the variance of the non-technology shocks in order to reproduce Galí's impact coefficient on hours? We find that the variance of non-technology shocks must be large, large enough so that the SVARs do miserably in terms of bias and size of confidence bands.

(Note here that under the Sims-Cogley-Nason approach, whether or not Galí's DSVAR is misspecified is irrelevant. Galí's statistic is just a moment of the data that has happened to receive a lot of attention, with possibly no more interpretation than those in the standard Kydland and Prescott (1982) table of moments.)

A third answer to the question of reproducible features is that if the SVAR procedure works well, then the variance of the shocks should be consistent with the variance decompositions in the SVAR literature itself. Much of this literature, including Christiano, Eichenbaum, and Vigfusson (2003), attributes only a small fraction of the fluctuations to technology shocks. As CKM show, with the parameters set to generate any of these statistics, the SVAR responses are badly biased and have enormous confidence bands.

In sum, contrary to the argument of CEV, the U.S. data do not definitively show that technology shocks account for virtually all of the movements in output. Most of the literature agrees with us, including much of the previous work of CEV, both alone and in concert.

3.2 Does the Mixed OLS–Newey-West Estimator Uniformly Improve on OLS?

Perhaps the most interesting part of CEV's work is their proposed estimator of impulse responses with long-run restrictions. They argue that this estimator, which splices the OLS estimator and a Newey and West (1987) estimator, "virtually eliminates the bias" (CEV 2006, p. 3) associated with the standard OLS estimator and thus makes the errors of their approach tiny. In this sense, CEV argue that it does not matter whether technology shocks account for almost all of the fluctuations in output because their new estimator takes care of the bias problem.

We disagree. The results of Mertens (2006) show that actually the new estimator does not even uniformly improve on the standard OLS estimator. Unfortunately, the new estimator is thus not a comprehensive solution for the problems with long-run restrictions.

To understand these issues, use the notation of Mertens (2006) to write the standard OLS estimator of the impact coefficient matrix A_0 as

$$A_0^{OLS} = \left(I - \sum_i B_i^{OLS} \right) Chol(S_X(0)^{OLS}),$$

where B_i^{OLS} denotes the regression coefficient matrices from the VAR and $Chol(S_X(0)^{OLS})$ denotes the Cholesky decomposition of the OLS estimate of the spectral density matrix $S_X(0)^{OLS}$ of the variables in the VAR at frequency zero. Here

$$S_X(0)^{OLS} = \left(I - \sum_i B_i^{OLS} \right)^{-1} \Omega^{OLS} \left(I - \sum_i B_i^{OLS} \right)^{\prime -1},$$

where Ω^{OLS} is the OLS estimate of the covariance matrix of residuals from the VAR.

CEV propose replacing $S_X(0)^{OLS}$ with a spectral density estimator along the lines of Newey and West (1987), with a Bartlett weighting scheme given by

$$S_X(0)^{NW} = \sum_{k=-b}^{b} \left(1 - \frac{|k|}{b+1} \right) E_T X_t X_{t-k}',$$

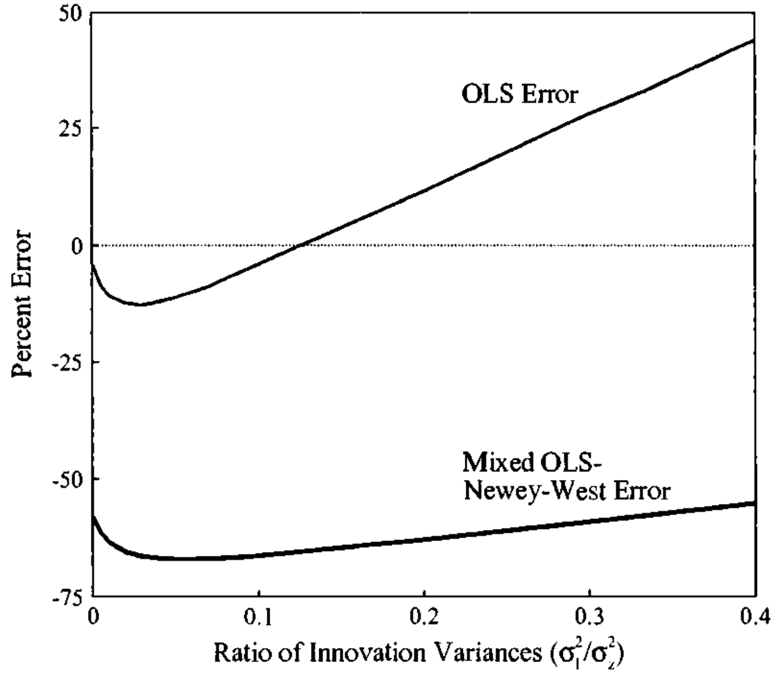
where X_t is the data, T is the sample length, E_T is the sample moments operator, and b is a truncation parameter.[5]

Figure 1.16, taken from Mertens (2006), displays the impact errors resulting from the use of the OLS estimator and the mixed OLS–Newey-West estimator, with four lags in the VAR, $b = 150$, $T = 180$, various values of σ_l^2/σ_z^2, and the rest of the parameters set as in CEV. The figure shows that when non-technology shocks are small, the CEV estimator has a larger bias than does the OLS estimator. As Mertens shows, if non-technology shocks are large enough, the positions eventually reverse. Clearly, the mixed OLS–Newey-West estimator is not uniformly better than the OLS estimator. (For more details, see Mertens (2006).)

3.3 *Are DSVARs Misspecified?*

It is somewhat of a puzzle to me why, in their broad assessment of SVARs, CEV focus on the LSVAR literature, which does not have eco-

Impact Errors Using the OLS and Mixed OLS–Newey-West Estimators



Source: Mertens (2006)

Figure 1.16
The Mixed OLS–Newey-West Estimator Is Not Uniformly Better

nomic results and has garnered neither much attention nor publications, and ignore the DSVAR literature, which both does and has. (See CKM's discussion of Fernald (2005), Gambetti (2006), and the LSVAR literature for details supporting this assertion.)

Both CKM and CEV argue that the DSVAR literature has a problem, but we disagree as to what it is. CKM argue that the only mistake in the DSVAR literature is that it uses the common approach rather than the Sims-Cogley-Nason approach; that is, this literature compares empirical SVARs to inappropriate objects in the model. In this comparison, the lag-truncation bias is severe enough that it flips the sign of the estimated impulse response. CEV argue that the DSVAR literature makes a different mistake. In particular, CEV argue that the procedure of differencing hours has "an avoidable specification error" (CEV, p. 26).

They seem to conclude that, even with an infinite amount of data, the DSVAR impulse responses will not coincide with the model's impulse responses. We disagree: CKM address the issue of misspecification directly and argue that the DSVAR procedure has no specification error of importance.[6]

CKM argue this result in two steps. The first step in our argument is that with a QDSVAR, with α close to 1, say, .99, Galí (1999) would have obtained impulse responses virtually indistinguishable from those he actually obtains in his DSVAR in which he sets α equal to 1. In this sense, for all practical purposes, we can think of Galí as having run a QDSVAR. The second step in our argument is that, for any $\alpha < 1$ and a long enough data set, the QDSVAR will get exactly the right answer. That is, with the lag length chosen to be suitably increasing with sample size, the sample impulse responses in the QDSVAR procedure will converge in the relevant sense to the model's impulse response; that is, $\bar{A}_i\,(p = \infty, T = \infty) = D_i(\theta)$. In this precise sense, contrary to what CEV claim, this procedure has no specification error of importance.

Marcet (2005) shows something subtler. He shows that with the DSVAR procedure in which α equals 1, the sample impulse responses from a procedure in which the lag length increases appropriately with sample size converge in the relevant sense to the model's impulse response. Marcet notes that his Proposition 1 seems to directly contradict the, at least implicit, claims of Christiano, Eichenbaum, and Vigfusson (2003).[7]

So, with large samples, researchers have no a priori reason to prefer the LSVAR procedure to the QDSVAR procedure, and with α close to 1 in samples of length typical to that in postwar data, the QDSVAR is indistinguishable from the DSVAR. Beyond that, small-sample issues do lead one specification to be preferred. Quasi-differencing lessens the amount of Hurwicz-type small-sample bias in estimating the parameters of a highly correlated series like per capita hours. Thus, at least a priori, the QDSVAR seems to be preferable to the LSVAR.

Nevertheless, the QDSVAR turns out to actually do worse than the LSVAR. When CKM decompose the mean impulse response error into small-sample bias and lag-truncation bias, we find that even though the QDSVAR has smaller Hurwicz-type bias, it has a much larger lag-truncation bias for reasonable parameters; the QDSVAR does worse. That is a quantitative result, however. We are not sure that it holds in a large class of models with a wide variety of parameters.

3.4 *Does Adding More Variables to the SVARs Help?*

CEV argue that, even though for a wide variety of circumstances, SVARs with long-run restrictions are uninformative, they can be informative in special cases—for example, when more variables are added to an LSVAR. Contrary to the impression we get from CEV, there is no disagreement on this point. Indeed, part of the point of CKM is to prove analytically exactly what the special cases are.

A commonly cited example of an economy in which SVARs with long-run restrictions work well is Fisher's (2006) model (see Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005)). CKM show that Fisher's model is a special case of our Proposition 4, a case when LSVARs can be informative. In this sense, we obviously agree with CEV about the validity of our proposition. We do not think, however, that an approach that works only in special cases has much to offer researchers seeking a reliable, generally applicable tool.

4 The Common Approach with Short-Run Restrictions

The use of the common approach on SVARs with long-run restrictions thus has little to recommend it. What about using it on SVARs with short-run restrictions? CEV claim that with this type of SVAR, their approach is a state-of-the-art technique that is useful for guiding theory. They focus on short-run restrictions that are satisfied in models which satisfy certain timing assumptions, often referred to as *recursive* assumptions. CEV claim to show that when a model satisfies such an assumption, SVARs with short-run restrictions perform remarkably well in small samples. And CEV imply that, because of this finding, this technique can be used broadly to distinguish promising models from the rest.

Since the CKM work has nothing to do with short-run restrictions, I have not studied the details of CEV's claims about how well the short-run restrictions work in practice with small samples and therefore have nothing to disagree with on these small-sample claims. Nevertheless, I do disagree with CEV's main message with respect to short-run restrictions in this area. As other researchers do, CEV ignore two critiques which seem to be widely thought of as devastating for much of the literature that uses SVARs with short-run restrictions. These critiques are of a theoretical nature, not about some problems with small samples. These critiques thus imply that, regardless of how well the short-run restrictions work with small samples, they are of little value in guiding the development of a broad class of monetary research. Hence, these critiques need to be addressed with a precise theoretical argument, not with some small-sample results.

The main critique of the SVAR literature with short-run restrictions is the *Lucas and Stokey critique* of Lucas and Stokey (1987). The point of this critique is that the particular class of short-run identifying assumptions made by CEV and related work in the short-run SVAR literature do not apply to a broad class of models and hence are of little use in guiding the development of a broad class of research.

The upshot of this critique is that some of the prominent researchers in the short-run SVAR literature have drastically overreached the conclusions of their studies. The short-run identifying assumptions in their work apply to only a tiny subset of monetary models, but the SVAR results have been used to rule out models not in that tiny subset. This mismatch between assumptions and models is a serious problem for this work.

A simple way for researchers in the short-run literature using the common approach to inoculate themselves from the Lucas and Stokey critique is to include in an appendix a list of the papers in the existing literature that satisfy their proposed identifying assumptions. Unfortunately, for most of the identifying schemes that I have seen, that list would be extremely short and would exclude most of the famous monetary models that constitute the core of theoretical monetary economics. If researchers are able to invent new identifying schemes for which this list is both broad and long, then this literature would have a much greater impact on guiding the development of monetary theory than it currently does. Doing so would constitute progress.

To understand my claim that the current literature is subject to the Lucas and Stokey critique, consider the recursiveness assumption itself. As Christiano, Eichenbaum, and Evans (1998, p. 68) explain, "The economic content of the recursiveness assumption is that the time t variables in the Fed's information set do not respond to the time t realizations of the monetary policy shock." To see how this assumption might be satisfied in a model, note that if the monetary authority at time t sets its policy as a function of time t variables, including output, consumption, and investment, as it does in Christiano, Eichenbaum, and Evans (2005), then the model must have peculiar timing assumptions in which, in a quarterly model, after a monetary shock is realized, private agents cannot adjust their output, consumption, and investment decisions during the remainder of the quarter. Whether or not one agrees that this timing assumption is peculiar, it is irrefutable that this timing assumption is not satisfied in the primary models in the monetary literature. I illustrate this point in figure 1.17, which lists the typical classes of models studied in monetary economics. (Technically, for all the models in the large rectangle, the impulse responses from the SVAR procedure do not converge in the relevant sense to the impulse responses in the model, so that $\bar{A}_i\,(p = \infty, T = \infty) \neq D_i(\theta)$.)

As an illustration of the claim that some of the short-run literature overreaches, consider the exposition by Christiano and Eichenbaum (1999) of the research agenda of the monetary SVAR literature. This exposition draws on the well-cited comprehensive survey by

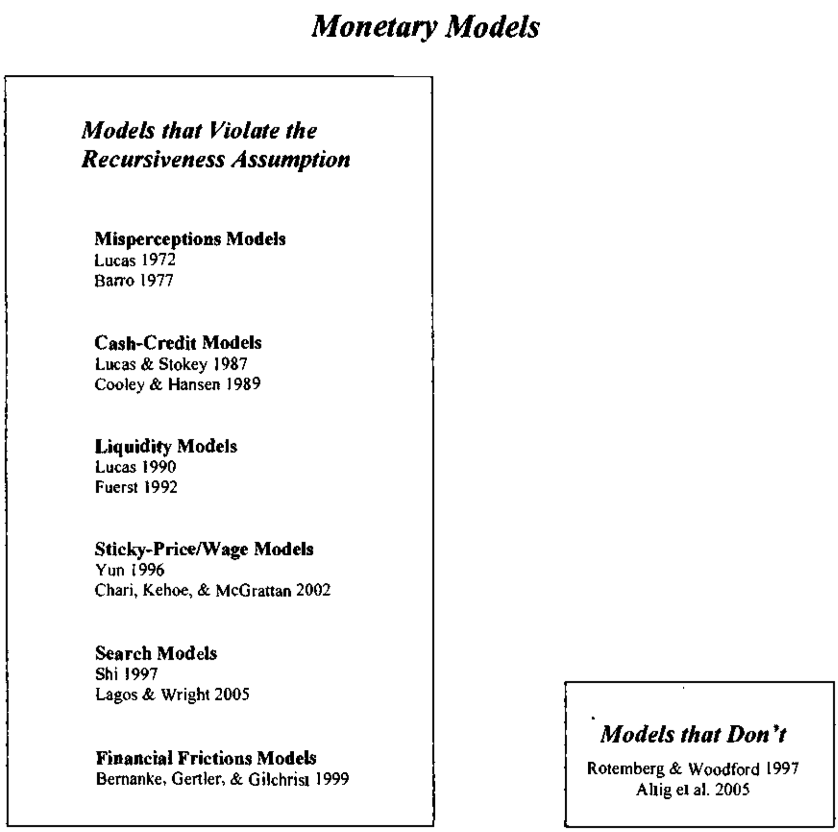


Figure 1.17
CEV's Recursiveness Assumption Does Not Apply to Most Monetary Models
(Representative Classes of Existing Monetary Models, Grouped Whether They Violate or Satisfy CEV's Recursiveness Assumption)

Christiano, Eichenbaum, and Evans (1998) of the short-run SVAR literature, which is the clearest statement of the research agenda of the monetary SVAR literature that I could find.

Christiano and Eichenbaum (1999) start with a summary of the so-called facts and a brief note that some identifying assumptions have been used to establish them:

In a series of papers, we have argued that the key consequences of a contractionary monetary policy shock are as follows: (i) interest rates, unemployment and inventories rise; (ii) real wages fall, though by a small amount; (iii) the price level falls by a small amount, after a substantial delay; (iv) there is a persistent decline in profits and the growth rate of various monetary aggregates; (v) there is a hump-shaped decline in consumption and output; and (vi) the US exchange rate appreciates and there is an increase in the differential between US and foreign interest rates. See CEE [Christiano, Eichenbaum, and Evans] (1998) for a discussion of the literature and the role of identifying assumptions that lie at the core of these claims.

Christiano and Eichenbaum (1999) then go on to reject some models that, they say, are not consistent with those facts. In particular, based on their SVAR-established facts, they reject both Lucas' (1972) island model and Lucas' (1990) liquidity model. These claims are clearly overreaching. Since Lucas' two models in particular do not satisfy the peculiar timing assumptions needed to justify the recursive identifying assumption in the SVAR, how is it logically coherent to reject those models based on the SVAR-established facts?

A potential objection to the Lucas and Stokey critique is that the SVAR literature is not overreaching because some of the models that violate the recursiveness assumption satisfy some other identifying assumptions that researchers have made, and for these other assumptions, SVAR researchers have found similar qualitative patterns. The second main critique of the short-run literature, the *Uhlig critique* of Uhlig (2005), dismisses this objection. The Uhlig critique is that the atheoretical SVAR specification searches are circular: "the literature just gets out what has been stuck in, albeit more polished and with numbers attached" (Uhlig 2005, p. 383). Uhlig argues that the reason the other identifying assumptions find similar answers is that the answers are essentially built into the search algorithm. Uhlig suggests that the algorithm used to find some SVAR results is, perhaps unconsciously, to pick a pattern of qualitative results and then do an atheoretical search over patterns of zeros, lists of variables to include, and periods of time to study, so that the resulting SVAR impulse responses reproduce the desired qualitative results. If this description is accurate, then I am sympathetic with Uhlig's conclusion that not much is to be learned from this branch of the short-run SVAR literature.

Note, again, that neither of these critiques would apply if, when comparing models and data, researchers simply followed the Sims-Cogley-Nason approach. Under that approach, the issue of whether the identifying assumptions of an SVAR hold in a model doesn't come up. The impulse responses from the SVAR on the data simply define some sample statistics that are coherently compared to the analogous statistics from the model. That is, now letting $A_i^{us}(p = 4, T = 180)$ and $\bar{A}_i\ (p = 4, T = 180 \mid \theta)$ denote the impulse responses obtained from an SVAR with short-run restrictions, using standard (simulated) method-of-moments logic, it makes perfect sense to compare these two even though $\bar{A}_i\ (p = \infty, T = \infty) \neq D_i(\theta)$.

In sum, if the SVAR literature with short-run restrictions followed the research agenda advocated by Sims (1989) and applied by Cogley and Nason (1995), then it would be on firm statistical and logical grounds.

5 Concluding Remarks

Let me be clear about what I am advocating in practice. For researchers willing to make a quantitative comparison between a model and the data, all I am advocating basically is changing several lines of computer code—replacing the theoretical impulse responses, $D_i(\theta)$, with the more relevant empirical responses derived from applying the SVAR procedure to the model, $\bar{A}_i\ (p = 4, T = 180 \mid \theta)$, in the relevant spots where the comparison between model and data is being made. For researchers who just want to run SVARs in the data and chat about what it means for a model, all I am advocating is a change in claims. Replace the claim about having robustly discovered what happens after a particular type of shock with a more precise claim about having documented what type of impulse responses should arise in a model when an SVAR with 4 lags and 180 observations is run on the data from it. Changing these several lines of code or text will vastly increase the intellectual impact of the approach.

It is puzzling to me that CEV and CKM seem to agree on two of the three key facts; yet we somehow disagree on their primary implication.

We agree on these two facts about the common approach:

- The common approach sometimes makes large errors relative to the Sims-Cogley-Nason approach. In particular, with long-run restrictions, SVARs do miserably unless technology shocks account for virtually all of the fluctuations in output.

- The common approach sometimes excludes most models of interest while the Sims-Cogley-Nason approach does not. For example, with short-run restrictions, the recursive identifying assumptions apply to only a tiny subset of the existing monetary models.

We disagree on one significant fact about interpreting the U.S. data:

- CEV argue that the evidence definitively implies that technology shocks account for virtually all of the fluctuations in output. CKM argue that while one can find statistics supporting this view, 20 years of macroeconomic research and some simple statistics show that shocks involving something other than technology play a sizable role in generating fluctuations in output and other variables.

And we disagree on the overriding implication of these facts:

- CEV argue that the common approach is a state-of-the-art technique that can be saved with a mechanical fix and analyst restraint:

  –For the long-run restriction branch of the SVAR literature, CEV argue that a mixed OLS–Newey-West estimator essentially eliminates the errors of the common approach.

  –For the short-run restriction branch, CEV think that, in order to avoid the over-reaching of some of the recent work in the area, researchers should be much more careful to delineate exactly to which work they claim their analyses apply. (This view is implicit in their conference discussions of early drafts of CKM's work.)

- CKM argue, to the contrary, that the common approach is not a state-of-the-art technique and that it should be abandoned in favor of one that is. The Sims-Cogley-Nason approach has firm statistical and theoretical foundations and thus avoids the type of statistical and logical errors of the common approach.

Beyond the specifics of the debate between CEV and CKM, my bottom line is simple: Let's stop worrying about the size of the errors in the old SVAR literature and instead start moving forward with the more promising Sims-Cogley-Nason approach that has potential to help us advance theory.

Acknowledgments

This work is a response to the comments of Lawrence Christiano, Martin Eichenbaum, and Robert Vigfusson on the critique of structural VARs with long-run restrictions by V. V. Chari, Patrick Kehoe, and Ellen McGrattan. For very helpful comments, the author thanks his coauthors as well as numerous others, including his coauthors, Tim Cogley, Jesus Fernandez-Villaverde, Bob Hall, Chris House, Narayana Kocherlakota, Ricardo Lagos, Monika Piazzesi, Juan Rubio-Ramirez, Tom Sargent, and Jim Stock. The author also thanks the NSF for financial support and Kathy Rolfe and Joan Gieseke for excellent editorial assistance. Any views expressed here are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Endnotes

1. The idea of the Sims-Cogley-Nason approach is to compare the exact small-sample distribution of the estimator from the model (with short lags) to the small-sample estimate (with short lags) from the data. This is a contrast to the common approach, which makes no attempt to deal with any of the issues that arise with a small sample. At a logical level, as long as the small-sample distribution is approximated well, either by hand, which is exceedingly difficult, or by a computer, which is easy, the Sims-Cogley-Nason approach seems to clearly dominate the common approach.

2. Note that, at least for the environment considered by CKM, since the Hurwicz-type small-sample bias is small, the comparison of

$A_i^{US}(p = 4, T = 180)$ to $\bar{A}_i\,(p = 4, T = \infty \mid \theta)$

would eliminate most of the error in the common approach and would allow the researcher to use standard asymptotic formulas. We view this comparison as a rough-and-ready approximation to the one in the Sims-Cogley-Nason approach.

3. This reasoning helps explain why the bulk of the real business cycle literature has not adopted the unit root specification. In this sense, technically, the SVAR results really have little to say about this literature. But that point has already been forcefully made by McGrattan (2005).

4. Note that, as other parameters in the model shift, so does the size of σ_l^2 needed to produce a certain volatility in hours. In this sense, whether or not a certain value of σ_l^2 is small or not should be judged by whether or not the model with this parameter can produce the observed volatility of hours in the data.

5. The spectral density at frequency zero is defined as $S_X(0) = \Sigma_{k=-\infty}^{\infty} EX_tX'_{t-k}$. The estimator of Newey and West (1987) is a truncated version of this sum that replaces population moments with sample moments and weights these sample moments to ensure positive definiteness.

6. In an interesting recent paper, Dupaigne, Fève, and Matheron (2006) take a different approach to show that while the DSVAR performs poorly in the laboratory of an economic model, this poor performance has nothing to do with specification error. These authors consider an economic model with nonstationary hours so that there is clearly no "avoidable specification error" in the DSVAR. Nonetheless, they show that DSVARs perform poorly in short samples for exactly the same reasons that DSVARs perform poorly in the CKM analysis. Dupaigne, Fève, and Matheron go on to show that the estimation procedures of structural models based on the common approach—in which parameters are chosen to minimize the distance between the theoretical impulse responses and the SVAR impulse responses—lead to systematically biased estimates. These authors argue that the natural estimation procedure based on the Sims-Cogley-Nason approach resoundingly dominates the procedure based on the common approach.

7. Part of the disagreement in this regard may come from a failure to precisely distinguish between two types of noninvertibility problems. The type considered by Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) are nontrivial and difficult to deal with without using a detailed economic theory. As Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) discuss and Marcet (2005) and CKM show, however, the type of knife-edge invertibility issues that come from differencing a stationary series are much more trivial and are easy to deal with.

References

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde. 2005. "Firm-Specific Capital, Nominal Rigidities, and the Business Cycle." NBER Working Paper no. 11034. Cambridge, MA: National Bureau of Economic Research.

Barro, R. 1977. "Unanticipated Money Growth and Unemployment in the United States." *American Economic Review* 67(2): 101–115.

Bernanke, B. S., M. Gertler, and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In J. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, Vol. 1C. Amsterdam: North-Holland.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan. 2002. "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?" *Review of Economic Studies* 69(3): 533–563.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan. 2006. "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theories?" Research Department Staff Report 364, Federal Reserve Bank of Minneapolis.

Christiano, L. J., and M. Eichenbaum. 1999. "The Research Agenda: Larry Christiano and Martin Eichenbaum Write about Their Current Research Program on the Monetary Transmission Mechanism." *EconomicDynamics Newsletter* 1(1). Society for Economic Dynamics, available at <http://www.economicdynamics.org/>.

Christiano, L. J., M. Eichenbaum, and C. L. Evans. 1998. "Monetary Policy Shocks: What Have We Learned and to What End?" NBER Working Paper no. 6400. Cambridge, MA: National Bureau of Economic Research.

Christiano, L. J., M. Eichenbaum, and C. L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113(1): 1–45.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson. 2003. "What Happens after a Technology Shock?" NBER Working Paper no. 9819. Cambridge, MA: National Bureau of Economic Research.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson. 2006. "Assessing Structural VARs." NBER Working Paper no. 12353. Cambridge, MA: National Bureau of Economic Research.

Cogley, T., and J. M. Nason. 1995. "Output Dynamics in Real-Business-Cycle Models." *American Economic Review* 85(3): 492–511.

Cooley, T. F., and G. D. Hansen. 1989. "The Inflation Tax in a Real Business Cycle Model." *American Economic Review* 79(4): 733–748.

Dupaigne, M., P. Fève, and J. Matheron. 2006. "Avoiding Pitfalls in Using Structural VARs to Estimate Economic Models." Banque de France. Mimeo.

Fernald, J. G. 2005. "Trend Breaks, Long-run Restrictions, and the Contractionary Effects of Technology Improvements." Working Paper no. 2005-21. Federal Reserve Bank of San Francisco.

Fernandez-Villaverde, J., J. F. Rubio-Ramirez, and T. J. Sargent. 2005. "A, B, C's (and D's) for Understanding VARs." NBER Technical Working Paper no. 0308. Cambridge, MA: National Bureau of Economic Research.

Fisher, J. D. M. 2006. "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." *Journal of Political Economy* 114(3): 413–451.

Francis, N., and V. A. Ramey. 2005. "Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited." *Journal of Monetary Economics* 52(8): 1379–1399.

Fuerst, T. S. 1992. "Liquidity, Loanable Funds, and Real Activity." *Journal of Monetary Economics* 29(1): 3–24.

Galí, J. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89(1): 249–271.

Galí, J., and P. Rabanal. 2005. "Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data?" In M. Gertler and K. Rogoff, eds., *NBER Macroeconomics Annual 2004*. Cambridge, MA: MIT Press, 225–288.

Gambetti, L. 2006. "Technology Shocks and the Response of Hours Worked: Time-Varying Dynamics Matter." Universitat Pompeu Fabra. Mimeo.

Kydland, F. E., and E. C. Prescott. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica* 50(6): 1345–1370.

Lagos, R., and R. Wright. 2005. "A Unified Framework for Monetary Theory and Policy Analysis." *Journal of Political Economy* 113(3): 463–484.

Lucas, R. E., Jr. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4(2): 103–124.

Lucas, R. E., Jr. 1990. "Liquidity and Interest Rates." *Journal of Economic Theory* 50(2): 237–264.

Lucas, R. E., Jr., and N. L. Stokey. 1987. "Money and Interest in a Cash-in-Advance Economy." *Econometrica* 55(3): 491–513.

Marcet, A. 2005. "Overdifferencing VAR's Is OK." Universitat Pompeu Fabra. Mimeo.

McGrattan, E. R. 2005. "Comment on Galí and Rabanal's 'Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data?' " In M. Gertler and K. Rogoff, eds., *NBER Macroeconomics Annual 2004*. Cambridge, MA: MIT Press, 289–308.

Mertens, E. 2006. "Shocks to the Long Run: A Note on Mixing VAR Methods with Non-Parametric Methods." Study Center Gerzensee. Mimeo.

Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3): 703–708.

Rotemberg, J. J., and M. Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." In B. S. Bernanke and J. J. Rotemberg, eds., *NBER Macroeconomics Annual 1997*. Cambridge, MA: MIT Press, 297–346.

Shi, Shouyong. 1997. "A Divisible Search Model of Fiat Money." *Econometrica* 67(1): 75–102.

Sims, C. 1989. "Models and Their Uses." *American Journal of Agricultural Economics* 71(2): 489–494.

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

Yun, T. 1996. "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles." *Journal of Monetary Economics* 37(2): 345–370.

Comment

Mark W. Watson, Princeton University and NBER

1 Introduction

Sometimes structural VARs work. Sometimes they don't. That is, in some situations SVARs can be used for reliable statistical inference about structural features of the economy, while in other situations SVARs provide misleading inference. Whether or not a SVAR will work depends on the structure of the economy and on the particulars of the SVAR.

There are three situations in which a SVAR will not work. First, a SVAR will not work if it is based on faulty identification restrictions. For example, VAR analysis based on an incorrect Wold causal ordering of the errors will lead to faulty inference. This is widely understood. Second, a SVAR will not work if the structural shocks under study cannot be recovered from current and past values of the variables used in the VAR. This is the "invertibility problem" discussed in the context of VARs in Hansen and Sargent (1991, 2007), Lippi and Reichlin (1994), and elsewhere. Third, a SVAR will not work when the data contain little information about key parameters. Said differently, a SVAR is a system of linear simultaneous equations and inference may be affected by unreliable or "weak" instruments.

The paper by Christiano, Eichenbaum, and Vigfusson (CEV) concerns the third problematic situation for SVARs. They use appropriately identified SVARs to study invertible model economies, thereby eliminating concerns about the first two problems. In their model economies, they vary the numerical value of parameters, and this changes the amount of information in realizations from the models. This allows them to determine the validity of statistical inferences based on SVAR analysis for a range of the model's parameter values. In the context of the models considered (a simple RBC model and a sticky-price variant), they determine a range of values for the model parameters for which

SVARs provide reliable inference about the dynamic effect of technology shocks on employment, and a range of values of the model parameters for which SVAR inference is unreliable. They argue that the U.S. economy is characterized by parameter values for which SVAR inference is reliable for the purpose of determining the effect of technology shocks on employment.

While CEV provide a systematic numerical study of the question and provide some intuition for their results, it is interesting to push a little harder on the question of *why* their SVARs fail. Constructively, it is also interesting to ask whether, in the context of the models used by CEV, the reliability of a proposed SVAR can be diagnosed using standard statistical procedures. The comment focuses on these two questions. The next section outlines a simplified version of the CEV model, and the following section uses this simplified model to explain why the SVAR worked for some parameter values, why it failed for others, and how (in principle at least) this could have been determined by a statistical test.

2 A Simplified Version of the Simple RBC Model

The two-shock RBC model discussed in section 2 of CEV leads to the following equations for labor productivity, y_t/l_t, and employment, l_t

$$\Delta\ln(y_t/l_t) = \gamma_y - \alpha\Delta\ln(l_t) + (1-\alpha)\ln(z_t) + \alpha\Delta\ln(k_t) \tag{1}$$

$$\ln(l_t) = \gamma_l + a_l\tau_{l,t} + a_z\ln(z_t) + \tilde{a}_z\ln(z_{t-1}) + a_k\ln(\hat{k}_t) \tag{2}$$

where (2) includes both $\ln(z_t)$ and $\ln(z_{t-1})$ to incorporate both the "Standard" and "Recursive" form of the model used by CEV. The simplified version of the model analyzed here suppressed the constant terms and the terms involving the capital stock. Constants play no role in the analysis; capital is more important because it affects the dynamics, but with one exception discussed below, has no important affect on the econometric features of the model that I will discuss.

Using CEV's AR(1) specification for $\tau_{l,t}$, $\tau_{l,t} = \rho_l\tau_{l,t-1} + \sigma_l\varepsilon_t^l$, a straightforward calculation shows that, after eliminating the constants and capital, (1)–(2) can be rewritten as the VAR

$$\Delta\ln(y_t/l_t) = \beta_0\ln(l_t) + \beta_1\ln(l_{t-1}) + \eta_t \tag{3}$$

$$\ln(l_t) = \phi\ln(l_{t-1}) + \gamma_1[\Delta\ln(y_{t-1}/l_{t-1}) - \alpha\Delta\ln(l_{t-1})] \qquad (4)$$
$$+ \gamma_2[\Delta\ln(y_{t-2}/l_{t-2}) - \alpha\Delta\ln(l_{t-2})] + \upsilon_t$$

where the VAR coefficients are $\beta_0 = -\beta_1 = -\alpha$, $\phi = \rho_l$, $\gamma_1 = (\tilde{a}_z - a_z\rho_l)/(1 - \alpha)$, and $\gamma_2 = -\tilde{a}_z\rho_l/(1 - \alpha)$; the VAR errors are $\eta_t = (1 - \alpha)\sigma_z\varepsilon_t^z$ and $\upsilon_t = a_l\sigma_l\varepsilon_t^l + a_z\sigma_z\varepsilon_t^z$.

3 Estimation and Inference in the SVAR

CEV are interested in estimating the dynamic effect of the technology shock on employment. This would be easy if technology shocks were observed. They can be constructed from (3) (up to scale) from the observed data if the coefficients β_0 and β_1 were known. Thus, I will focus on estimating these coefficients. CEV discuss two restrictions that identify these coefficients: a short-run restriction and long-run restriction. I consider each in turn.

CEV's "short-run" restriction is that $a_z = 0$ in (2). This implies that the VAR shock in the employment equation does not depend on the technology shock. That is, $\upsilon_t = a_l\sigma_l\varepsilon_t^l$ in equation (4). This restriction implies that $\ln(l_t)$ is uncorrelated with η_t in (3), so that the coefficients in (3) can be estimated by OLS. Standard theory shows that OLS performs well in regressions involving stationary variables. Consistent with this, CEV find the SVARs perform well using this short-run restriction. (A closer look at the CEV results suggests some deterioration in the quality of SVAR for longer IRF lags. This is probably caused by the high persistence in l_t (more on this below); with persistent regressors VAR impulse response estimators are nicely behaved (approximately normal) for short lags but have non-normal distributions at longer lags for reasons discussed in Sims, Stock, and Watson (1990).)

CEV's "long-run" restriction is that long-run movements in $\Delta\ln(y_t/l_t)$ come solely from the (scaled) productivity shock η_t. With $|\phi| < 1$, this equivalent to the restriction that $\beta_0 = -\beta_1$ in (3). Imposing this restriction, (3) becomes

$$\Delta\ln(y_t/l_t) = \beta\Delta\ln(l_t) + \eta_t \qquad (5)$$

where $\beta = \beta_0 = -\beta_1$. Equation (5) cannot be estimated by OLS because $\Delta\ln(l_t)$ is correlated with η_t. It can be estimated by instrumental variables if valid instruments can be found. Candidates are provided by equation (4), which implies $\Delta\ln(l_t)$ can be expressed as

$$\Delta\ln(l_t) = [\phi - 1] \times \ln(l_{t-1}) + \gamma_1[\Delta\ln(y_{t-1}/l_{t-1}) - \alpha\Delta\ln(l_{t-1})] \qquad (6)$$
$$+ \gamma_2[\Delta\ln(y_{t-2}/l_{t-2}) - \alpha\Delta\ln(l_{t-2})] + v_t.$$

Thus the variables $\ln(l_{t-1})$, $[\Delta\ln(y_{t-1}/l_{t-1}) - \alpha\Delta\ln(l_{t-1})]$, and $[\Delta\ln(y_{t-2}/l_{t-2}) - \alpha\Delta\ln(l_{t-2})]$ are candidate instruments. Because these variables are dated $t - 1$ and earlier, they are uncorrelated with η_t, so they satisfy the "orthogonality" condition for valid instruments. But a valid instrument must also be "relevant," which means that it must be correlated with $\Delta\ln(l_t)$. From (6), these instruments are valid if at least one of the coefficients $[\phi - 1]$, γ_1, or γ_2 are nonzero. If all of these coefficients are zero, the instruments are not valid. If at least one of the coefficients is non-zero and "large," standard theory suggests that the IV estimator will perform well. If the coefficients are non-zero but "small," then the instruments are "weak" in the sense of Staiger and Stock (1997), and IV estimator will perform poorly. (Pagan and Robertson (1998), Cooley and Dwyer (1998), and Sarte (1997) discuss the weak instrument problem in a model like this.) Evidently, the performance of the SVAR hinges on the values of ϕ, γ_1, and γ_2.

In their two-shock models, CEV use two values of ϕ (= ρ_l): 0.986 in the "CEV" parameterization and 0.952 in the "CKM" parameterization. In both cases $\phi - 1$ is close to zero, suggesting a weak instrument problem associated with $\ln(l_{t-1})$. The size of γ_1 and γ_2 in (6) are governed by the size of a_z and $\tilde{a}_z$ in (2). These parameters govern the size of the effect of the technology shock on employment. When this effect is small, γ_1 and γ_2 are small, $[\Delta\ln(y_{t-1}/l_{t-1}) - \alpha\Delta\ln(l_{t-1})]$ and $[\Delta\ln(y_{t-2}/l_{t-2}) - \alpha\Delta\ln(l_{t-2})]$ are weak instruments, and the IV estimator will perform poorly. This explains why CEV found that the SVAR model performed poorly when the technology shock explained a small fraction of the variance of employment.

In summary, in the context of the models and questions discussed in CEV, standard SVAR analysis will be reliable when strong instruments can be found for $\Delta\ln(l_t)$, but will be unreliable when only weak instruments are available. Of course, instrument relevance is something that can be checked in the data. For example, Staiger and Stock (1997) suggest that an F-statistic less than ten in the "first-stage" regression is an indication of a potential weak instrument problem. As CEV note, several papers have used long-run identified SVARs and post-war quarterly data to estimate the effect of technology shocks on employment. A leading example is CEV (2003) which uses a version of (3)–(4) that includes constants and additional lags. The first stage F-statistic for that model ranges from 9 to 11 depending on the details of the specifications (number of lags, sample period, and so forth). Thus, based on the Staiger-Stock rule of thumb ($F > 10$), there is cause for some (perhaps slight) concern. But, the Staiger-Stock analysis uses stationary regressors, and l_t is highly persistent in the U.S. data. My suspicion, although the details have not been worked out, is that the Staiger-Stock rule of thumb is too small in this case. Thus, weak instruments may well be a problem in SVAR specifications like those used in CEV (2003). (Interestingly, weak instruments do not seem to be a problem when the SVAR is specified using Δl_t in place of l_t, as in Galí (1999) and Francis and Ramey (2005). The first stage F for these SVARs is greater than 30.)

Finally, it is useful to discuss two other interesting findings in CEV: That the long-run SVAR performs well when it utilizes knowledge of the true value of the zero-frequency spectrum of $\Delta\ln(y_t/l_t)$ and $\ln(l_t)$, and that some of these gains can be achieved using a non-VAR based estimator of the zero-frequency spectrum. To see why the zero-frequency spectrum helps, consider an extreme case in which $\gamma_1 = \gamma_2 = 0$ in (6), so that $\ln(l_{t-1})$ is the only potential instrument. In this case, the IV estimator of β in (5) is $\hat{\beta}^{IV} = \hat{\pi}^{OLS}/(\hat{\phi} - 1)$ where $\hat{\pi}^{OLS}$ is the OLS estimator from the regression of $\Delta\ln(y_t/l_t)$ onto $\ln(l_{t-1})$ and $\hat{\phi}$ is an estimator of ϕ. Because $\phi - 1$ is close to zero, small sampling error in $\hat{\phi}$ leads to large (and non-Gaussian) sampling error in $\hat{\beta}^{IV}$. (This is another way of characterizing the weak-instrument problem.) If the value of ϕ was known, then the problem would be eliminated, and if sampling error in $\hat{\phi}$ could be reduced, then the problem would be mitigated. Not surprisingly the zero-frequency spectrum of the series provides a lot of information about ϕ, which is incorporated in the SVAR when the spectrum is known. This explains the good performance of the SVAR that uses the true value of the zero-frequency spectrum. The performance of the SVAR that uses the non-VAR estimator of the zero-frequency spectrum is somewhat more mysterious. My guess is that something like the following is going on: When capital is included in the model, the data are described by a VARMA model, so that the VAR needs a large number of lags to adequately capture the model's long-run dynamics. This leads to truncation bias in the estimated value of ϕ computed using a short-lag VAR, and this truncation bias is eliminated using the alternative estimator proposed by CEV. Analyzing the properties of this SVAR estimator would be interesting and non-standard because it relies on an inconsistent estimator of the zero-frequency spectrum. (The estimator used by CEV uses an untruncated Bartlett kernel.) Kiefer and Vogelsang (2002) and Müller (2005) discuss the usefulness of this inconsistent estimator in other contexts.

References

Christiano, L. J., M. Eichenbaum, and R. Vigfusson. 2003. "What Happens After a Technology Shock." Manuscript, Northwestern University.

Cooley, T. F., and M. Dwyer. 1998. "Business Cycle Analysis Without Much Theory: A Look at Structual VARs." *Journal of Econometrics* 83: 57–88.

Francis, N., and V. A. Ramey. 2005. "Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregrate Fluctuations Revisited." *Journal of Monetary Economics* 52: 1379–1399.

Galí, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89: 249–271.

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

Hansen, L. P., and T. J. Sargent. 2007, forthcoming. *Recursive Linear Models of Dynamic Economies. Princeton, NJ:* Princeton University Press.

Kiefer, N., and T. Vogelsang. 2002. "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation." *Econometrica* 70: 2093–2095.

Lippi, M., and L. Reichlin. 1994. "VAR Analysis, Nonfundamental Representations, Blashke Matrices." *Journal of Econometrics* 63(1): 307–325.

Müller, U. K. 2005. "A Theory of Long-Run Variance Estimation." Manuscript, Princeton University.

Pagan, A. R., and J. Robertson. 1998. "Structural Models of the Liquidity Effect." *Review of Economics and Statistics* 80: 202–217.

Sarte, Pierre. 1997. "On the Identification of Structural Vector Autoregressions." *Federal Reserve Bank of Richmond Economic Quarterly* 83(3): 45–67.

Sims, C. A., J. H. Stock, and M. W. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots" (with C. A. Sims and J. H. Stock). *Econometrica* 58(1).

Staiger, D., and J. H. Stock. 1997. "Instrumental Variable Regression with Weak Instruments." *Econometrica* 65(3): 557–586.

Discussion

Lawrence Christiano responded to a number of points made by Ellen McGrattan in her presentation of Patrick Kehoe's comment. First, he disagreed with the view that the appropriate way to conduct the line of research in which he and his coauthors were engaged was to run a VAR on the data from the model, and then run the same VAR on data from the actual economy, and compare the two. While he said that he was generally sympathetic to this approach, he thought that this was not the appropriate approach for the problem they were studying. In their case, they were trying to assess how well a VAR estimator is able to estimate a particular feature of a model, namely how hours respond to a productivity shock. He thought that for this sampling theory question, the approach they used was the appropriate one.

Christiano agreed with McGrattan that the conclusions of VAR analysis using first differenced data could be very misleading. He, however, felt that the fact that VARs are sensitive to first differencing was not a reason to throw out VARs. He pointed out that many other statistical procedures, such as correlations, are sensitive to first differencing, but this has not led economists to discard these other procedures.

Christiano then commented that it was true that in the context of RBC models, the VARs they estimate with long run restrictions tend to produce very large confidence intervals. He said that this was due to the fact that sampling uncertainty is big in data coming from RBC models. He stressed that VARs are good in that they will correctly tell the researcher that the sampling uncertainty is large and that there is not much information in the data generated by the RBC model. Christiano then stressed that VARs work very well with short run restrictions and that the VAR literature that relies on short run restrictions has had a large impact on how macroeconomists think about business cycles.

Christiano and Christopher Sims both questioned Kehoe's dismissal of short run restrictions as not being implied by rational expectations models. Christiano remarked that an equilibrium condition in an economic model will imply zero restrictions in a VAR whenever the VAR includes more variables than does the equilibrium condition. Sims remarked that in a paper with Tao Zha, he had analyzed a DSGE model and shown that a short run identifying restriction of the type usually used in the VAR literature was consistent with this model.

Edward Prescott remarked that there are many exciting and interesting puzzles in macroeconomics on issues of great importance, such as the fact that labor supply in Europe is depressed by 30 percent and the fact that Japan has lost a decade of growth due to low productivity growth. In light of these facts, he felt it was unfortunate that the discussion in this session seemed to be about how many angels can dance on the head of a pin.

Prescott also remarked that the Lucas critique had taught economists that estimating structural VARs is inconsistent with dynamic economic theory. Sims responded that he felt it was great to get input on statistical methods from Real Business Cycle theorists.

Sims commented that it was always something of a mystery why researchers should expect an exact match between the number of structural shocks and the number of variables they happened to include in their VAR. He noted that in a paper with Zha, he had shown that it was not in fact necessary to have an exact match of this kind, and that in their model, the monetary policy shocks were well identified even though the system was not invertible.

Sims wondered why Chari, Kehoe, and McGrattan had turned their original critique of Galí and Rabanal into a critique in which they claimed that SVARs were no good in general. He wanted to be sure that they were not trying to argue against all probability based inference because in his opinion much of the discussion had that flavor. Sims suggested that if Chari, Kehoe, and McGrattan were going to conclude that SVARs were no good in general that they should suggest an alternative methodology. Chari replied that in their paper on Business Cycle Accounting they had advanced one alternative procedure and that more generally there were many kinds of state space procedures that could provide an alternative to VARs. Chari noted that a drawback of all these alternative procedures was that they relied somewhat more heavily on assumptions about the structure of the economy. He felt, however, that the minimalist approach embodied in SVARs seemed not to be very useful.

Sims remarked that the weak instruments problem Mark Watson had discussed in his comment was easily dealt with by using likelihood based inference, even in the case when the instruments are highly autocorrelated.

Chari felt that the SVAR procedure as originally envisioned by Blanchard and Quah and later applied by many other authors was "totally cool stuff." This opinion was based on the fact that the SVAR literature came up with strong, clear, and confident results that held the promise of allowing researchers to reject certain classes of models and focus on other classes. He then explained that all he, Kehoe, and McGrattan had wanted to do in their paper was to subject SVARs to a simple test. If they generated data from a model where they knew what the response of hours to a productivity shock was, would a SVAR be able to identify that the data came from the model? He noted that the examples in their paper did not raise questions about VARs in general, but rather only attempted to assess how good SVARs are at identifying whether data is generated from the particular model they specified. Their findings were that when demand shocks are important, then the SVAR does not perform particularly well.

Chari then remarked that different papers in the SVAR literature analyzing the same question had found very different results based on seemingly small differences in the variables being used. In reply, Martin Eichenbaum disagreed that the differences in the data were small. He furthermore noted that when the sample was restricted to a more recent sample period where the differences in the data were in fact small, the differences in results disappeared.