Federal Reserve Bank of Minneapolis
Research Department

A Critique of Structural VARs Using Real Business Cycle Theory*

V. V. Chari, Patrick J. Kehoe, Ellen R. McGrattan

Working Paper 631

Revised February 2005

ABSTRACT

The main substantive finding of the recent structural vector autoregression literature with a differenced specification of hours (DSVAR) is that technology shocks lead to a fall in hours. Researchers have used these results to argue that standard business cycle models in which technology shocks leads to a rise in hours should be discarded. We evaluate the DSVAR approach by asking the following: Is the specification derived from this approach misspecified when the data is generated by the very model the literature is trying to discard, namely the standard business cycle model? We find that it is misspecified. Moreover, this misspecification is so great that it leads to mistaken inferences that are quantitatively large. We show that the other popular specification which uses the level of hours (LSVAR) is also misspecified with respect to the standard business cycle model. We argue that an alternative approach, the business cycle accounting approach, is a more fruitful technique for guiding the development of business cycle theory.

*Chari, University of Minnesota and Federal Reserve Bank of Minneapolis; Kehoe, Federal Reserve Bank of Minneapolis and University of Minnesota; McGrattan, Federal Reserve Bank of Minneapolis and University of Minnesota. The authors thank the National Science Foundation for support.

The goal of the Structural Vector Autoregression (SVAR) approach is to identify promising classes of business cycle models using a simple time series technique. The approach has two popular specifications both of which use data on labor productivity and hours. The differenced specification, called the DSVAR, uses the first difference in hours while the level specification, called the LSVAR, uses the level of hours. We evaluate the SVAR procedure under both specifications by applying it to data generated from a standard business cycle model and find that both specifications are misspecified. With respect to the DSVAR our key finding is that the misspecification leads to quantitatively large mistaken inferences about standard business cycle models. With respect to the LSVAR our key finding is that in samples as long as those for postwar U.S. data the misspecification leads to uninformative inferences while for much longer samples the misspecification leads to mistaken inferences.

The SVAR approach begins with the idea that it is possible to obtain impulse responses from the data using only a minimal amount of economic theory.[1] These impulse response functions are the responses of the model's economic system to innovations in various shocks. The hope is that the SVAR assumptions nest most business cycle models, at least approximately, so that the impulse responses obtained from the VAR set the standard for the theory: any promising model must produce impulse responses similar to those from the VARs.

We focus on the DSVAR and the LSVAR literatures that study what happens after a technology shock. The findings of the two literatures are quite different. The main finding of the DSVAR literature is that a technology shock leads to a fall in labor input. Gali (1999), Francis and Ramey (2003), and Gali and Rabanal (2004) use the DSVAR procedure to draw the inference that this evidence dooms existing real business cycle models as unpromising and points to other models, such as sticky price models, as a more promising class of models. For example, Francis and Ramey (2003, p.2) say, "...the original technology-driven real business

[1] See, among others, Shapiro and Watson 1988, Blanchard and Quah 1989, Gali 1999, Francis and Ramey 2003, Christiano, Eichenbaum and Vigfusson 2003, Uhlig 2003, and Gali and Rabanal 2004.

cycle hypothesis does appear to be dead." Likewise, Gali and Rabanal (2004, conclusion) state, "The bulk of the evidence reported in the present paper raises serious doubts about the importance of changes in aggregate technology as a significant (or, even more, a dominant) force behind business cycles ..."

In the LSVAR literature the range of results reported by researchers is very wide. Francis and Ramey (2004) argue that the LSVAR evidence shows that real business cycle models are dead. Christiano, Eichenbaum and Vigfusson (2003) argue that their LSVAR results imply that these models are alive and well. Gali and Rabanal (2004) argue that their LSVAR results, by themselves, are inconclusive. As we document below, while all three studies use slightly different methodologies, their sharply contrasting results are driven almost entirely by differences in the underlying data. It is worth noting that all three of these studies use very similar conceptual measures of productivity and Christiano, Eichenbaum and Vigfusson and Gali and Rabanal use very similar conceptual measures of hours. That the LSVAR results are so sensitive to seemingly minor differences in measuring productivity and hours raises doubts about the reliability of the LSVAR procedure for drawing inferences about underlying models.

Both branches of the SVAR literature make several assumptions to identify the underlying shocks, often labelled as *demand* and *technology* shocks. This literature views two identifying assumptions as key: (i) demand shocks have no permanent effect on the level of labor productivity while technology shocks do and, (ii) the demand and technology shocks are orthogonal.

We test the SVAR procedure as follows. We generate data from an economic model, apply the SVAR procedures to this data and ask whether the impulse responses identified by the SVAR procedures are close to the model's impulse responses. To tilt our test in favor of the SVAR procedure we focus mainly on a very stripped-down business cycle model, referred to as the *baseline* model. This model satisfies the two key identifying assumptions of the

SVAR literature.

Consider first our implementation of the DSVAR procedure. One critique of this procedure is that in all economic models hours per person is bounded and therefore the stochastic process for hours per person cannot literally have a unit root. Hence the DSVAR procedure is misspecified with respect to *all* economic models and thus is useless for distinguishing between broad classes of models. In our view this critique is simplistic. We are sympathetic with the view expressed in the DSVAR literature that the unit root specification is best viewed as a statistical approximation for variables with high serial correlation. (See, for example, Francis and Ramey 2003, p.6 for an eloquent defense of this position.)

We follow the literature in adopting a three step procedure. First, we conduct unit root tests on hours to the data generated from an estimated version of the baseline model and retain those sequences for which tests cannot reject a unit root. Second, we apply the Akaike criterion to the sequences retained in the first step and retain those sequences for which the criterion picks a lag length less than or equal to 4. We then estimate the DSVAR impulse responses for each of these retained sequences using the standard identifying assumptions in the DSVAR literature.

When we apply this procedure to data generated from the model of the same length as post–war U.S. data, we conclude that this shock leads to a decline in hours. This conclusion is mistaken because in our model an innovation to technology leads to a rise in hours. Thus, we show that a researcher who applied the DSVAR procedure to our model would systematically draw the wrong inference.

We then attempt to deconstruct the source of the mistaken inference with respect to the impulse responses. We focus on an auxiliary assumption which is not emphasized as central in the SVAR literature: the stochastic process for the first differences of labor productivity and hours is well approximated by an autoregressive representation with a small number of lags. We show that this auxiliary assumption fails to be satisfied in a standard real business cycle model and that this failure is at the heart of the mistaken inference. We show that in our model the autoregressive representation has an infinite order in which the coefficients die off very slowly. We go on to show that the key reason for these features of the autoregressive representation is that our model has capital. Since capital accumulation is central to business cycle models these features of the autoregressive representation are likely to show up in most models. We also show that given data from our model as short as postwar data standard lag length tests do not detect the need for more lags.

Consider next the LSVAR procedure. In our two step implementation of this procedure, we first apply the Akaike criterion to data generated from our baseline model and retain those sequences for which the criterion picks a lag length less than or equal to 4. In our second step we apply an LSVAR to the retained sequences, and compare the resulting impulse responses to the model's impulse responses. When we apply this SVAR procedure to data from our baseline model of the same length as post–war U.S. data, the resulting confidence bands on the impulse responses are so wide that the procedure is essentially uninformative. For example, the impulse responses of most real business cycle models as well as most sticky price models are well within these bands. In this sense, this specification sheds no light on the models it is meant to illuminate.

We also apply the LSVAR procedure to a data series many times longer than post–war U.S. data and find that the procedure leads to mistaken inference. We argue that the source of the mistaken inference is similar to that of the DSVAR test: the stochastic process for the productivity and hours is not well approximated by an autoregressive representation with a small number of lags. Moreover, standard lag length tests do not detect the need for more lags.

Note that in our test we use data generated from an economic model. We do so because in the model we can take a clear stand on what constitutes a technology shock. Hence, the question of whether fluctuations in factor productivity in U.S. data come from changes in technology or from other forces is totally irrelevant for our test. We emphasize that our test is a logical analysis of the inferences drawn from the SVAR procedures and neither asks nor depends on why factor productivity in the U.S. data fluctuates.

In this paper, we argue that even the most standard business cycle model fails to satisfy the assumptions underlying the SVAR approach. Hence, the representations derived from the SVAR procedure are misspecified with respect to the very models the procedure is meant to shed light on. Of course, in almost all applied work the statistical representations of the theory are, in a narrow sense, technically misspecified with respect to the class of theoretical models under consideration. Our contribution is to show that this misspecification is quantitatively large.

Finally, we note that our critique builds on those in a number of papers that we discuss below, especially Hansen and Sargent (1980, 1991) and Cooley and Dwyer (1998).

1. The Structural VAR Procedure

We briefly review a version of the Blanchard and Quah (1989) structural VAR procedure recently used by Gali (1999), Gali and Rabanal (2004), Francis and Ramey (2003).

The procedure starts with a vector autoregression of the form

$$(1) \qquad X_t = B_1 X_{t-1} + \ldots + B_p X_{t-p} + v_t$$

where the error terms have variance-covariance matrix $Ev_t v_t' = \Omega$ and are orthogonal at all leads and lags so that $Ev_t v_s' = 0$ for $s < t$. The vector X_t is given by $(x_{1t}, x_{2t})'$ where $x_{1t} = \Delta \log(y_t/l_t)$ is the first difference of the log of labor productivity and x_{2t} is a measure of the labor input. We consider two different specifications: in the difference specification (DSVAR), x_{2t} is the first difference in the log of the labor input, in the level specification (LSVAR), x_{2t} is the log of the labor input.

This vector autoregression, as it stands, should be thought of as a reduced form of an economic model. Specifically, the error terms v_t have no structural interpretation. It is convenient to invert this vector autoregression in order to express it in its equivalent moving average form

$$(2) \quad X_t = v_t + C_1 v_{t-1} + C_2 v_{t-2} + \ldots$$

where $C_1 = B_1, C_2 = B_1 C_1 + B_2, C_3 = B_1 C_2 + B_2 C_1 + B_3$ and so on.

The idea behind the structural VAR procedure is to use the bare minimum of economic theory together with the reduced form model (2) to back out structural shocks and the responses to those shocks. To that end, consider the following structural model which links the variations in the log of labor productivity and the labor input to a (possibly infinite) distributed lag of two shocks, thought of as a technology shock and a demand shock. The empirical model is given by

$$(3) \quad X_t = A_0 \varepsilon_t + A_1 \varepsilon_{t-1} + A_2 \varepsilon_{t-2} + \ldots$$

where $\varepsilon_t = (\varepsilon_t^z, \varepsilon_t^d)'$ represent the technology and demand shocks with $E\varepsilon_t \varepsilon_t^{'} = \Sigma$ and $E\varepsilon_t \varepsilon_s^{'} = 0$ for $s \neq t$.

The structural parameters A_i and Σ are related to the reduced form parameters C_i and Ω by $A_0 \Sigma A_0' = \Omega$ and $A_j = C_j A_0$ for $j \geq 1$. The structural shocks ε_t are related to the reduced form shocks v_t by $A_0 \varepsilon_t = v_t$ so that $\varepsilon_t = A_0^{-1} v_t$.

The assumptions used to identify the structural parameters from the reduced form parameters are as follows. The first assumption is that technology shocks and demand shocks are orthogonal. If we interpret the structural shocks as having been scaled by their standard deviations, we can express this assumption as $\Sigma = I$ so that $E\varepsilon_t \varepsilon_t' = I$. The second assumption, the long run restriction, is that

$$(4) \quad \sum_{i=0}^{\infty} A_i(1,2) = 0$$

where $A_i(1,2)$ is the element in the first row and second column of the matrix A_i. This assumption is meant to capture the idea that demand shocks cannot affect the level of labor productivity in the very long run, while technology shocks may be able to do so.

These two assumptions give four equations in the four unknowns of A_0 that allow us to solve for A_0 in terms of the estimated parameters C_i and Ω. The first assumption, $\Sigma = I$, gives three nonlinear equations from $A_0 A_0' = \Omega = (\omega_{ij})$, namely, $A_0(1,1)^2 + A_0\,(1,2)^2 = \omega_{11}$, $A_0(1,1)A_0(2,1) + A_0(1,2)A_0(2,2) = \omega_{12}$, $A_0(2,1)^2 + A_0(2,2)^2 = \omega_{22}$. Let $\bar{C} = \sum_{j=0}^{\infty} C_j$ be the sum of the moving average coefficient where $C_0 = I$. The second assumption, (4), then gives one linear equation, namely

$$\bar{C}(1,1)A_0(1,2) + \bar{C}(1,2)A_0(2,2) = 0.$$

The rest of the structural parameters A_j, $j \geq 1$ follow from $A_j = C_j A_0$ for $j \geq 1$. Under the sign conventions that $A_0(1,1) > 0$, so that productivity rises with a technology shock, and $A_0(2,2) < 0$, so that hours falls with a demand shock, the solution to these equations is

$$(5) \quad A_0(2,2) = -\sqrt{\lambda}, A_0(1,2) = f A_0(2,2), A_0(1,1) = \sqrt{\omega_{11} - f^2\lambda}, A_0(2,1) = \frac{\omega_{12} - f\lambda}{A_0(1,1)}$$

where $f = -\bar{C}(1,2)/\bar{C}(1,1)$ and $\lambda = \left[\omega_{11}\omega_{22} - (\omega_{12})^2\right] / \left[\omega_{11} - 2f\omega_{12} + f^2\omega_{22}\right]$. Note, for later use, that $A_0(2,1)$ is the response of hours on impact to a technology shock.

Here we focus on the differenced specification. We estimate the bivariate VAR using ordinary least squares on U.S. quarterly data from 1959:1–2003:4. As is typical in the SVAR literature we use four lags in the VAR. (See, for example, Gali 1999.) In terms of the data, for output we use real GDP per capita from NIPA and for the labor input we use CPS data on total hours worked in the noninstitutional population over sixteen divided by the total noninstitutional population over sixteen[2].

In Figure 1 we report the impulse response the log level of hours obtained from (3) along with 95 percent confidence bands computed using a bootstrap Monte Carlo procedure with 1000 replications. We find that a positive shock to technology, ε_t^z, leads to a persistent and statistically significant fall in hours. These findings are similar to those in the literature obtained using other measures of labor productivity and labor input, including those in Gali (1999), Gali and Rabanal (2004), and Francis and Ramey (2003 and 2004). All of these papers use conceptually similar measures of productivity and hours. Francis and Ramey (2004) adjust the hours series to account for demographic changes.

2. A Standard Business Cycle Model

Our test uses versions of an entirely standard business cycle model. (See, among a host of others, McGrattan 1989.) We consider a stripped down version of the model, called the *baseline* model, that meets the two key identifying assumptions of the SVAR procedure, namely that demand shocks have no permanent effect on the level of labor productivity and that the demand and technology shocks are orthogonal. The baseline model is the best case scenario for the SVAR procedure. This model has two stochastic variables, technology shocks Z_t, which have a unit root, and an orthogonal tax on labor τ_{lt}. It also has a constant investment tax τ_x and a constant level of normalized government spending $\bar{g} = g_t/Z_t$.

Consumers maximize expected utility over per capita consumption c_t and per capita labor l_t, $E_0 \sum_{t=0}^{\infty} [\beta(1+\gamma)]^t U(c_t, l_t)$ subject to the budget constraint

$$(6) \quad c_t + (1+\tau_x)[(1+\gamma)k_{t+1} - (1-\delta)k_t] = (1-\tau_{lt})w_t l_t + r_t k_t + T_t$$

[2] The labor data was kindly provided to us by Ueberfeldt and Prescott.

where k_t denotes the per capita capital stock, w_t the wage rate, r_t the rental rate on capital, β the discount factor, γ the growth rate of population, δ the depreciation rate of capital, and T_t lump-sum taxes.

The firms' production function is $F(k_t, Z_t l_t)$, where Z_t is labor-augmenting technical progress. Firms maximize $F(k_t, Z_t l_t) - r_t k_t - w_t l_t$. The resource constraint is

$$(7) \quad c_t + g_t + (1+\gamma)k_{t+1} = y_t + (1-\delta)k_t$$

where y_t and g_t denote per capita output and per capita government consumption.

In the baseline model the stochastic process for the two shocks $\log Z_t$ and τ_{lt} is

$$(8) \quad \log Z_{t+1} = \mu_z + \log Z_t + \sigma_z \varepsilon^z_{mt+1}$$

$$(9) \quad \tau_{lt+1} = (1-\rho)\bar{\tau}_l + \rho\tau_{lt} + \sigma_l \varepsilon^d_{mt+1}$$

where ε^z_{mt} and ε^d_{mt} are standard normal random variables that we refer to as the technology and demand shocks for the model. These variables are independent of each other and i.i.d. over time. We refer to $\sigma_z \varepsilon^z_{mt}$ and $\sigma_l \varepsilon^d_{mt}$ as the innovations to technology and labor. These innovations have standard deviations σ_z and σ_l. The constant μ_z is the drift term in the random walk for technology. The constant $\bar{\tau}_l$ is the mean of the labor tax and ρ is the persistence parameter for the labor tax. We restrict ρ to be less than one so that innovations to the labor tax have no permanent effects.

We use functional forms and parameter values familiar from the business cycle literature. We assume that the production function has the form $F(k,l) = k^\theta l^{1-\theta}$ and the utility function has the form $U(c,l) = \log c + \psi \log(\bar{l} - l)$. We choose the capital share $\theta = .35$, the depreciation rate $\delta = .046$, and the discount factor $\beta = .97$. The population growth rate is 1.5 percent at an annual rate, the time allocation parameter $\psi = 2.24$, and the endowment of time $\bar{l} = 5,000$ hours per year.

In estimating the stochastic processes for our model we use no direct data on tax rates. Rather, the tax rate on labor and investment are treated as unobserved and are inferred from the model's equilibrium conditions. We think of the tax on labor as standing in for a variety of distortions that affect the static first order condition for labor. Likewise, we think of the tax on investment as standing in for a variety of distortions to the intertemporal Euler equation. (See Chari, Kehoe and McGrattan 2003 for a formalization of these ideas.)

We use a standard Kalman filtering approach to estimate our model. To use this approach we write a log-linearized version of our model in state space form in terms of stationary variables. To do so we let $\hat{k}_t = k_t/Z_{t-1}$ and $z_t = Z_t/Z_{t-1}$ and $s_t = (\log z_t, \tau_{lt})$. The state in period t is $S_t = (\log \hat{k}_t, s_t, \log \hat{k}_{t-1}, s_{t-1}, 1)$ denote the state in period t. The state's evolution is determined by the capital stock decision rule

$$\log \hat{k}_{t+1} = \gamma_0 + \gamma_k \log \hat{k}_t + \gamma_z \log z_t + \gamma_l \tau_{lt}$$

and exogenous process for the shocks $s_t = (\log z_t, \tau_{lt})$ which in matrix form is given by

$$(10) \quad s_{t+1} = \bar{P} + P s_t + Q \varepsilon_{mt+1}.$$

where $\bar{P} = (\mu_z, (1-\rho)\bar{\tau}_l)$ and P is a matrix with all zeros except the lower right element $P_{22} = \rho$, Q is a diagonal matrix with diagonal elements (σ_z, σ_l) and $\varepsilon_{mt} = (\varepsilon^z_{mt}, \varepsilon^d_{mt})$. We can stack up these equations to give the state transition equation

$$(11) \quad S_{t+1} = F S_t + G \varepsilon_{mt+1}.$$

Let the observable variables $Y_t = (\Delta \log y_t, \log l_t, \Delta \log x_t, \Delta \log c_t)'$ denote the growth rate of output, the log level of labor, the growth rate of investment, and the growth rate of consumption. The decision rule for labor in the model has the form

$$\text{(12)} \quad \log l_t = \phi_0 + \phi_k \log \hat{k}_t + \phi_z \log z_t + \phi_l \tau_{lt}.$$

and the decision rule for the growth rate of output has a similar form. We can write the observed variables $Y_t = HS_t$ where H is a matrix of coefficients of the linear decision rules for the vector Y_t. Because we have only two shocks in our model we add a small amount of measurement error u_t to what is observed and write the observed variables as

$$\text{(13)} \quad Y_t = HS_t + u_t.$$

In practice, we choose the parameters of the covariance matrix of the measurement error to allow just enough measurement error to avoid invertibility problems during estimation.

We then use the maximum likelihood procedure described in McGrattan (1989) and Anderson et al. (1996) to estimate the parameters $\bar{P}$, P, and Q of the vector AR1 process using data on output, the labor input, investment, and consumption. We report the parameter values for the stochastic processes for the baseline model in Table 1.

3. Testing the DSVAR Procedure

Here we simulate data from our baseline model and ask if the SVAR procedure using the differenced specification, or simply the *DSVAR procedure,* produces impulse responses similar to the theoretical impulse responses in the model. We find that it does not. In particular, a researcher who applied the DSVAR procedure to data from our model would conclude that in the underlying model a positive technology shock leads to a fall in hours on impact. In the underlying model, however, a positive technology shock leads to a rise in hours. Our main finding is that the DSVAR procedure systematically draws the wrong inferences about properties of the underlying model.

We begin with the impulse response from the baseline model to a technology shock. We start at a steady state and set the technology shock innovations $\varepsilon^z_{m0} > 0$, $\varepsilon^z_{mt} = 0$ for $t \geq 1$, and the labor tax innovations $\varepsilon^d_{mt} = 0$ for all t. In Figure 2 we see that on impact, a shock to technology leads to a persistent increase in hours. The left vertical axis follows the SVAR literature and measures the response to a one standard deviation shock in ε^z_{mt} while the right vertical axis follows the business cycle literature and converts this response to a one percent shock to total factor productivity. In order to relate the two vertical axes, note that 1 unit on the left vertical axis corresponds to $1/(1-\theta)\sigma_z$ units on the right axis. All our impulse response plots use two axes of this form. We see that on impact a one percent shock to total factor productivity leads to an increase in hours of .44 percent.

Consider now applying the DSVAR procedure to data generated by our baseline model. As we have discussed, we implement the DSVAR procedure in three steps. For any data series generated from our model we first conduct an Augmented Dickey–Fuller unit root test on hours (with a trend and 4 lags). If the test cannot reject a unit root we then apply the Akaike criterion to the sequences retained in the first step and retain those sequences for which the criterion picks a lag length less than or equal to 4. In our third step, we run the DSVAR[3].

One way to apply this procedure is as follows. Choose the sequence of shocks in the model to replicate the U.S. data on labor productivity and hours exactly. (This exact replication can be done because we have two shocks and two variables per period.) We then apply the SVAR procedure on this data generated by the model. By construction, an Augmented Dickey Fuller test will not reject a unit root in hours, since it did not for U.S. data. We then run two ordinary least squares regressions each with four lags on the first difference of the log of both labor productivity and hours. Clearly, the DSVAR specification will reproduce Figure 1 exactly. Thus, the responses to shocks that this procedure identifies

[3]We also implemented two other procedures. In one we simply replaced the ADF unit root test with other popular unit root tests. In the other we retained all data sequences, regardless of the results of the ADF test. Under both of these procedures our findings were virtually identical to the ones reported here.

as technology shocks is very different from the true responses to technology shocks in the model displayed in Figure 2. In particular, the DSVAR procedure gets the sign wrong on how technology shocks affect hours.

Our analysis so far leaves open the possibility that the sequence of shocks our model needs to produce the U.S. data is very unlikely. Perhaps, then, the DSVAR procedure fails for that atypical sequence of shocks but it would not fail for a more typical sequence. To address this possibility we treat the baseline model as the data-generating process and draw a large number of sequences of the same length as our data length, namely 180 quarters. We apply the Augmented Dickey Fuller test to the hours series for each sequence. We retain only those sequences for which the test cannot reject a unit root at the 5% level. The retained sequences represent about 80% of all the sequences. We then apply the Akaike criterion to the sequences retained in the first step and retain those sequences for which the criterion picks a lag length less than or equal to 4. We find that approximately 98% of the first step sequences are retained in this second step. (We also found similar retention rates applying likelihood ratio tests to 5 lags versus 4 lags.)

We report features of the impulse responses of hours to technology shocks that result from running the DSVAR specification on 1,000 retained sequences. In the left panel of Figure 3, we plot a histogram of the impact coefficients over the 1,000 sequences. The histogram shows that almost all of the impact coefficients are negative. The right panel of Figure 3 reports the range of estimated impulse responses over these 1,000 sequences for 12 periods. We construct this range by discarding the largest 2.5% and the smallest 2.5% of the impulse response coefficients in each period and report the range of the remaining 95% of the impulse response coefficients. In Figure 4 we plot the mean impulse response across these 1,000 sequences as well as the mean of the bootstrapped confidence bands across these same sequences. From Figures 2, 3, and 4, we see that the impulse responses from the DSVAR procedure are very different from the theoretical impulse responses.

We next ask, suppose for each of the 1,000 sequences a researcher tested the hypothesis that the impact coefficient of the DSVAR equals the theoretical impact coefficient at the 5% significance level. We find that such a researcher would mistakenly infer that the data did not come from our real business cycle model essentially 100% of the time.

Next we ask if our results are sensitive to the particular parameters of the stochastic process for shocks that we have used. In Figure 5 we report on the mean value of the impact coefficient and the mean value of the bootstrapped confidence intervals for various values of the relative volatility of the labor tax and the technology shock as measured by σ_l/σ_z for two values of the persistence parameter for labor taxes ρ. From Figure 5 we see that our key findings our not sensitive to values in this range.

There is a sense in which the range of parameter values we have considered contain most reasonable values. To see this note that we estimated the parameters of the baseline model's stochastic process using maximum likelihood on GDP and hours per person. Researchers in the SVAR literature often focus on productivity and hours measures for a subset of the economy–the business sector. In Table 1 we report on the parameters obtained when we reestimate the model on the data used by Gali (1999), Francis and Ramey (2004) and Christiano, Eichenbaum and Vigfusson (2003). All of these parameters fall within the range of those used in Figure 5.

4. Deconstructing the DSVAR Procedure's Mistaken Inferences

In addition to the main identifying assumptions the DSVAR procedure makes a key auxiliary assumption: in the models of interest $X_t = (\Delta(y_t/l_t),\ \Delta l_t)$ has an autoregressive representation that is well approximated with a small number of lags, typically 4. This assumption could fail for one of two reasons. First, such models do not have invertible moving average representations in X_t and therefore do not have an autoregressive representation. Second, the models do have invertible moving average representations but the associated autoregressive representations are not well approximated with a small number of lags. In this section we argue that lack of invertibility, while technically true, does not play an important role in the mistaken inference. We argue that instead the heart of the DSVAR's mistaken inference is that the autoregressive representation is not well approximated with a small number of lags. Moreover, standard lag length tests do not detect the need for more lags.

Technically, our model does not have an invertible moving average representation for $X_t = (\Delta(y_t/l_t), \Delta l_t)$. The reason is essentially that hours in our model do not have a unit root. For our model, specifying the SVAR in the difference in hours amounts to over–differencing hours and introduces a root of 1 in the moving average representation, which is at the edge of the noninvertibility region of roots. To show that this lack of invertibility is inessential in leading to the mistaken inference we consider a specification of the SVAR in which hours are quasi-differenced, referred to as the QDSVAR specification. In this specification $X_t = (\Delta(y_t/l_t), l_t - \alpha l_{t-1})$ where α is less than 1. With this specification it is easy to show that our model has an invertible moving average representation. Nevertheless, when α is close to 1 the impulse responses of the QDSVAR and the DSVAR are so close as to be indistinguishable. In what follows we will set the quasi-differencing parameter equal to .99. (It is worth noting that the literature contains several models in which the lack of invertibility of the moving average representation is not knife-edge. See, for example, Hansen and Sargent (1980) and Quah (1990).)

Let the invertible moving average representation of X_t be denoted

$$X_t = A_{m0}\varepsilon_{mt} + A_{m1}\varepsilon_{mt-1} + A_{m2}\varepsilon_{mt-2} + \ldots$$

where here and throughout the subscript m signifies the model. Recall that the covariance matrix of ε_{mt} is the identity matrix. This moving average representation can be written in the Wold form

$$(14) \quad X_t = v_{mt} + C_{m1}v_{mt-1} + C_{m2}v_{mt-2} + \dots$$

where $C_{mj} = A_{mj}(A_{m0})^{-1}$ and the shocks $v_{mt} = A_{m0}\varepsilon_{mt}$ have covariance matrix $\Omega_m = A_{m0}A'_{m0}$. Let $\bar{C}_m = \sum_{j=0}^{\infty} C_{mj}$. The autoregressive representation of X_t is

$$(15) \quad X_t = B_{m1}X_{t-1} + B_{m2}X_{t-2} + \dots + v_{mt}$$

where $B_{m1} = C_{m1}, B_{m2} = C_{m2} - B_{m1}C_{m1}, B_{m3} = C_{m3} - B_{m1}C_{m2} - B_{m2}C_{m1}$ and so on. The proof of the following proposition is available upon request.

Proposition 1. The autoregressive representation of X_t is infinite order and satisfies $B_{mj} = MB_{mj-1}$ for $j \geq 2$ where the matrix $M = C_{m2}(C_{m1})^{-1} - C_{m1}$ and has eigenvalues equal to the quasi-differencing parameter $\lambda_1 = \alpha$ and $\lambda_2 = (\gamma_k - \gamma_l\phi_k/\phi_l - \theta)/(1-\theta)$ where $\gamma_k, \gamma_l, \phi_k$ and ϕ_l are the coefficients of the decision rules and θ is the capital share.

Given our parameters (including the quasi-difference parameter $\alpha = .99$) the eigenvalues are $\lambda_1 = .99$ and $\lambda_2 = 97$. These large eigenvalues suggest that an autoregression with a small number of lags is a poor approximation to the infinite order autoregression. Of course, with many lags the autoregression approximates the infinite order autoregression well.

For an autoregression with one lag we can obtain an analytical expression for the impulse response coefficients. This autoregression is of the form

$$(16) \quad X_t = B_1X_{t-1} + v_t \text{ with } Ev_tv_t{}' = \Omega.$$

Let $\bar{C} = \sum_{j=0}^{\infty} C_j$ be the associated sum of the moving average coefficients. Recall that for an autoregression of the form (1) with an arbitrary number of lags the impact coefficients are given by (5). Inspecting (5) we see that to compute the impact coefficients we need only the estimated covariance matrix and the sum of the moving average coefficients. The proof of the following proposition is available upon request.

Proposition 2. For our model, the population estimate of the covariance matrix in (16) is

$$(17)\quad \Omega = \Omega_m + M\left(\Omega_m - \Omega_m V(X_t)^{-1}\Omega_m\right)M'$$

and the (inverse of the) sum of the moving average coefficients associated with (16) is

$$(18)\quad \bar{C}^{-1} = (\bar{C}_m)^{-1} + M(I-M)^{-1}C_{m1} + M(\Omega_m - V(X_t))V(X_t)^{-1}$$

where $V(X_t)$ is the covariance matrix of X_t.

Consider a researcher who incorrectly specifies that the autoregressive representation of X_t has only one lag and uses the SVAR procedure to uncover the associated impulse responses. This researcher would infer that the covariance matrix of innovations is Ω and that the sum of the moving average coefficients is $\bar{C}$. The expressions given in (5) can then be used to compute the impact coefficients. Since the Ω and $\bar{C}$ derived from the one lag VAR, in general differ from the model's Ω_m and $\bar{C}_m$, typically the impulse responses will differ as well. (Of course, if this researcher had specified an infinite number of lags the autoregression would have recovered the model's covariance matrix Ω_m, the sum of the model's moving average coefficients $\bar{C}_m$ and hence would have correctly uncovered the model's impulse responses.)

We next ask how many lags are needed in practice to obtain a good approximation. Obviously, in a short sample of length 180 it not possible to accurately estimate a QDSVAR with a large number of lags. We choose a sample length of 100,000 so that we can accurately estimate a QDSVAR with a large number of lags. In Figure 6a we display the model impulse response and the impulse response of the QDSVAR specification with 4 lags. Because of the long sample length the impulse response coefficients are precisely estimated and the associated bootstrapped confidence bands lie essentially on top of the impulse responses. Figure 6a shows that even with the long sample the QDSVAR with a lag length of 4 leads to systematically mistaken inference.

Furthermore, comparing the mean impulse response over the short samples in Figure 4 with the impulse response in the long sample in Figure 6a we see that the impact coefficients are roughly the same while the coefficients at longer lags differ somewhat. This comparison shows that the estimates of the impulse responses using short samples are biased, as is well–known in the literature on estimating autoregressions. The comparison in Figure 6a of the differences between the impulse responses of the model with that of the QDSVAR shows, however, that small sample biases are not the source of the mistaken inference.

In Figures 6b and 6c we display the QDSVAR responses for lag lengths p ranging from 4 to 200. Notice that even with 20 lags, because the confidence bands are narrow, the QDSVAR would lead to mistaken inference. From Figure 6b and 6c we note that the convergence to the model's impulse response function is not monotonic. Finally, note that we need over 200 lags for the QDSVAR to well–approximate the model's autoregressive representation.

We argue that the need for a large number of lags when we run the VAR with $X_t = (\Delta(y_t/l_t),\ l_t - \alpha l_{t-1})$ stems from the presence of capital in our model and the absence of capital in the VAR specification. We make this argument by showing that as we make capital less important by increasing its depreciation rate the impulse responses of the QDSVAR with only 4 lags converge to the model's impulse responses. In Figure 7, for two values of the depreciation rate we plot the impulse responses for the model and for the QDSVAR with 4 lags estimated on samples of length 100,000. We see that when the depreciation rate is 95%, so that capital essentially depreciates completely within a period, the model and the QDSVAR impulse responses are very close. Interestingly, even with depreciation rates of

20%, which is more than 4 times higher than our baseline value, the model and the QDSVAR impulse responses are not close.

It is worth noting that when the model has essentially one shock the QDSVAR impulse responses are close to the model impulse responses. Indeed in this case even a VAR with one lag produces impulse responses close to the one in the model. In the limit when the variance of the labor tax τ_l is zero, $M\Omega_m M' = M\Omega_m V(X_t)^{-1}\Omega_m M'$ so that from (17) we have that the covariance Ω in (16) is equal to that in the model Ω_m. Interestingly, because Ω_m is singular in the limit, the two impulse responses coincide even though the moving average coefficients from the one-lag specification C_j do not equal the model's moving average coefficients C_{mj}.

5. The LSVAR Procedure

Here we consider a popular alternative specification to the DSVAR we have discussed above the LSVAR, which replaces the first difference in the log of hours by the level of the log of hours. (See, Gali 1999, Francis and Ramey 2003, and Christiano, Eichenbaum and Vigfusson 2003 who debate whether the differenced or the level specification is the most appropriate.) We find that this procedure is also not a useful guide for developing business cycle models. In contrast to the DSVAR procedure which leads to systematically mistaken inferences, in samples as long as postwar data, the LSVAR procedure in unable to distinguish among broad classes of models. In particular, we argue that for such samples this procedure cannot distinguish between sticky price models and real business cycle models. Furthermore, in very long samples the procedure leads to systematically mistaken inferences.

A. Impulse Responses for 4 U.S. Datasets

We begin by displaying impulse responses and the associated confidence bands obtained by running the LSVAR specification with 4 lags using ordinary least squares on 4 U.S. datasets. It turns out that the impulse responses are very different for these four datasets. (It is worth noting that the impulse responses from the DSVAR specification on these same datasets are very similar.)

For the first series we use the data constructed by Francis and Ramey (2004) to estimate an LSVAR for the period 1948:1–2002:4. Their measure of productivity is the BLS series, "Index of output per hour, business." They construct a new measure of hours by adjusting the BLS series, "Index of hours in business" for government employment and for demographic changes. In Figure 8a we see that an innovation that results in a one percent increase in total factor productivity leads to a persistent decline in hours. On impact the decline is 1.9% and is significantly different from zero at the 5% level.

In the second we follow Christiano, Eichenbaum and Vigfusson (2003) who use the DRI economic database to estimate an LSVAR for the period 1948:1–2001:4. Their measure of productivity is business labor productivity (mnemonic LBOUT) and their measure of hours is business hours divided by civilian population over the age of 16 (mnemonics LBMN and P16). In Figure 8b we see that a positive technology shock leads to a persistent rise in hours. On impact a one percent increase in total factor productivity results in a .5% increase in hours. Notice that while the impact coefficient is not significantly different from zero, the response coefficients are significant from lag 3 onwards. (It worth pointing out that in their paper, Christiano, Eichenbaum and Vigfusson use an instrumental variables procedure proposed by Shapiro and Watson (1988) rather than our OLS procedure and they compute Bayesian confidence intervals rather than our bootstrapped confidence intervals. Comparing our Figure 8b with Figure 2 in their paper we see that we obtain essentially the same results that they do.)

For the third series we follow Gali and Rabanal (2004) and use data from 1948:1–2002:4. Their measure of productivity is business labor productivity constructed as the ratio of nonfarm business sector output to hours of all persons in the nonfarm business sector. For hours they use the ratio of nonfarm hours to civilian population over the age of 16. The source is Haver USECON database with mnemonics for output, hours, and population of

LXNFO, LXNFH and LNN. In Figure 8c, we see that a positive technology shock leads to a persistent but statistically insignificant rise in hours. On impact the rise is essentially zero and is not significantly different from zero at the 5% level.

In the fourth we start with a measure of output and a measure of the labor input using aggregate data from 1959:1–2003:4. We construct productivity as the ratio of output to hours. For output we use real GDP from NIPA adjusted for indirect business taxes. For both the productivity measure and the hours measure we use CPS data on total hours worked in the noninstitutional population over sixteen divided by the total noninstitutional population over sixteen. In Figure 8d a positive technology shock leads to a persistent but statistically insignificant rise in hours at all lags. Note that on impact the rise is essentially zero.

These sharply contrasting results lead researchers in the SVAR literature to draw sharply contrasting inferences. Francis and Ramey argue that this evidence shows the real business cycle models are dead. Christiano, Eichenbaum and Vigfusson argue that the models are alive and well. Gali and Rabanal argue that these results, by themselves, are inconclusive. They prefer the alternative DSVAR specification which they argue shows that real business cycle models are also dead.

One question that naturally arises is why in their papers do Christiano, Eichenbaum and Vigfusson (2004) and Gali and Rabanal (2004) get such different results for the LSVAR? In their paper, Christiano, Eichenbaum and Vigfusson (2004) focus on why the DSVAR and LSVAR for their data gives different impulse responses but do not mention that the impulse responses from their LSVAR are quite different from the LSVAR in Gali and Rabanal (2004). As we have shown above, the difference between the results in these papers lies in the difference in the underlying data and not in the method.

Interestingly, all three of these studies use very similar conceptual measures of productivity and Christiano, Eichenbaum and Vigfusson and Gali and Rabanal use very similar conceptual measures of hours. The sensitivity of the LSVAR results to seemingly minor differences in measuring productivity and hours raises some doubts about the reliability of the LSVAR procedure for drawing inferences about underlying models.

6. Testing the LSVAR Procedure

Our test of the LSVAR procedure has two steps. First, apply the Akaike criterion to data generated from our model and retain those sequences for which the Akaike criterion picks a lag length less than or equal to 4. Then, run the LSVAR specification with 4 lags on the retained sequences. (We found that for almost all sequences from our model standard lag length tests do not detect the need for more than 4 lags.)

As in the DSVAR test, we could, of course, choose the sequence of shocks in the model to replicate the series in the 4 datasets exactly. Clearly, the LSVAR specification will reproduce Figures 8a, 8b, 8c, and 8d exactly. Again, this analysis leaves open the possibility that the sequences of shocks our model needs to produce these datasets is very unlikely. To address this possibility we treat the baseline model as the data-generating process and draw a 1,000 sequences of the same length as our data length, namely 180 quarters. In the left panel of Figure 9 we display the histogram of the impact coefficients of the estimated LSVAR over the 1,000 sequences. The histogram shows that the distribution of the impact coefficients is bimodal with a substantial subset of the sequences having very negative coefficients. In the right panel of Figure 9 we plot the middle 95% range of impulse response coefficients. We see that the range of impact coefficients for a one percent shock to total factor productivity is huge: it ranges from -2.5% to 2%.

To get some understanding for why the histogram for the impact coefficient is bimodal recall from (5) that this coefficient is given by

$$(19) \quad A_0(2,1) = \frac{\omega_{12} - f\lambda}{\sqrt{\omega_{11} - f^2\lambda}}$$

where the parameters $\omega_{11}, \omega_{12}, f$ and λ are estimated from a VAR with 4 lags. Each simulation produces different values of these parameters. It turns out that typically the denominator of (19) is very small and that across runs the numerator switches signs. This feature of the simulations induces a bimodal distribution.

Notice just how large is the range over impact coefficients. Our real business cycle model has an impact coefficient of only .44 which is less than one–fifth of the upper end of this range. Notice that this range also includes –.44. That is, the range includes impact coefficients from models in which labor falls in response to a technology shock by as much as it rises in a real business cycle model. Actually, it includes models in which labor fall almost six times as much as it rises in a real business cycle model.

In Figure 10 we plot the mean impulse response and the mean of the bootstrapped confidence bands. It is clear that a wide range of theoretical impulse responses are consistent with the estimated impulse responses from the LSVAR.

We argue that this procedure is not useful for distinguishing sticky price models from flexible price models including real business cycle models. In sticky price models the responsiveness of hours to a technology shock depends on the extent to which the monetary policy accommodates the shock. For example, Gali, Lopez-Salido and Valles (2003) construct a simple sticky price model in which the monetary authority follows a Taylor rule and show that hours rise in response to a technology shock. They also show that if monetary policy is not all accommodative, hours fall in response to a technology shock. The range of responses for hours to technology shock in sticky price models is well within our 95% range.

We next ask, suppose for each of the 1,000 sequences a researcher tested the hypothesis that the impact coefficient of the LSVAR equaled the model's impact coefficient at the 5% significance level. We find that such a researcher would essentially never reject this hypothesis. We then ask, suppose the researcher tested the hypothesis that the impact coefficient of the LSVAR equaled the *negative* of the model's impact coefficient at the 5% significance level.

Such a researcher would also essentially never reject this hypothesis either. We interpret these findings as suggesting that at least with data of the kind that comes out of our model, the LSVAR is incapable of differentiating between models with starkly different impulse response functions.

A. Is Short Sample Length the Only Problem for the LSVAR Procedure?

The reason we find the LSVAR procedure useless for guiding theory is that the confidence bands contain the impulse responses of most models we want to discriminate between. Suppose then we consider a very long sample of length 100,000 so that the confidence bands are so narrow that they essentially lie right on top of the impulse response. Will the LSVAR procedure with 4 lags deliver an impulse response which is quite close to the model's impulse response? In Figure 11a we report on the impulse responses of an LSVAR with 4 lags for this long sample together with the model's impulse response. We see that the impulse response of the LSVAR is quite far from the model's impulse response[4]. In Figure 11b we show how the impulse responses from the LSVAR for lag lengths varying from $p = 4$ to 200. As with the DSVAR, we see that the impulse response from the LSVAR is a good approximation to the model only for extremely long lag lengths.

Some intuition for these results can be obtained from Propositions 1 and 2. Notice that these propositions hold for any quasi-differencing parameter $\alpha \in [0, 1)$. Hence they include the level SVAR in which $\alpha = 0$. From Proposition 1 we see that the autoregressive representation of $X_t = (\Delta(y_t/l_t),\, l_t)$ is infinite order with autoregressive coefficients that satisfy

$$(20) \quad B_{mj} = MB_{mj-1} \text{ for } j \geq 2$$

[4] It is worth noting that for the LSVAR the small sample bias is large. The size of this bias can be seen by comparing the estimates of the impulse response for the 4 lag specification for the short and the long data samples. Specifically, compare the mean impulse response in Figure 10 for the short sample to the impulse response for $p = 4$ in Figure 11 for the long sample. We see that these responses are very different. For example, in the long sample the impact coefficient is more than 3 times as large as that in the short sample.

where the matrix M has eigenvalues equal to $\lambda_1 = 0$ and $\lambda_2 = (\gamma_k - \gamma_l \phi_k/\phi_l - \theta)/(1-\theta)$. For our parameter values $\lambda_2 = .97$ and hence the autoregressive representation of X_t is poorly approximated with an autoregression with a small number of lags. Proposition 2 also applies so that typically the impulse responses of the LSVAR procedure are different from those of the model.

The reason the LSVAR needs such long lag lengths also seems to be driven by the role of capital. In that vein, in Figure 12 we show that only when the capital essentially depreciates completely every period does the LSVAR with 4 lags give a good approximation to the model's impulse response. An analytic way of seeing the importance of capital is to note in a version of the model without capital (obtained by setting the capital share parameter $\theta = 0$) the associated M in (20) is identically zero so that the autoregressive representation of X_t has only lag.

7. Adding Other Variables

So far we have focused on the SVARs with two variables: the log difference labor productivity and a measure of the labor input. In the SVAR literature, researchers often check how their results change when they include one or more extra variables in the SVAR. Here we investigate adding other variables in the SVARs and show that doing so leads to no significant change in our results.

Our earlier discussion of the SVAR specifications suggested that one of the problems with it was that it did not include a capital-like variable. In practice, researchers shy away from using measures of the level of the capital stock because they believe it is poorly measured and instead prefer variables such as the investment-output ratio. One conjecture is that including such a variable might diminish the need for estimating long lags in the SVARs. Hence, the short lag specification should yield accurate measures of the model's response to a technology shock. Here we show that the conjecture is incorrect.

To demonstrate this result we consider three variable SVARs in which we include as a third variable the log of the investment-output ratio x_t/y_t where $x_t = (1+\gamma)k_{t+1} - (1-\delta)k_t$. We also add a third shock by letting the tax on investment be stochastic and follow an autoregressive process

$$\tau_{xt+1} = (1-\rho_x)\bar{\tau}_x + \rho_x \tau_{xt} + \sigma_x \varepsilon^x_{mt+1}$$

where ε^x_{mt} together with our earlier shocks ε^z_{mt} and ε^d_{mt} are jointly normal, independent of each other and i.i.d. over time. To make our results parallel to our earlier ones with two shocks, we use the same parameters for the stochastic processes for the first two shocks as we did before. For the investment tax process we use $\rho_x = .90$ and $\sigma_x = .01$.

For brevity's sake we focus on the population predictions of the SVARs. In Figure 13, we compare the model's impulse response for labor against the population impulse response from a 4 lag LSVAR with two variables, $X_t = (\Delta(y_t/l_t),\ l_t)$ and one with three variables $X_t = (\Delta(y_t/l_t),\ l_t.x_t/y_t)$. We see that adding x_t/y_t does not improve the performance of the LSVAR. Indeed, for the values of ρ_x and σ_x we chose, it worsens the performance.

Some intuition for why adding a third variable does not significantly improve the SVARs performance is the following analogue of Proposition 1. Using the analogous notation to that proposition we then have the following.

Proposition 3. The autoregressive representation of X_t is infinite order and satisfies $B_{mj} = MB_{mj-1}$ for $j \geq 2$ where the matrix M has eigenvalues equal to the quasi-differencing parameter $\lambda_1 = \alpha$, $\lambda_2 = (1-\delta)/(1+g_y)$ and $\lambda_3 = 0$.

Here g_y is the growth rate of (total) real output. Given our parameters the eigenvalue λ_2 is .98. This large eigenvalue helps provide intuition for why an autoregression with a small number of lags is a poor approximation to the infinite order autoregression.

The conjecture we are considering is that including the investment-output ratio makes the LSVAR ($\alpha = 0$) with short lags yield a much better approximation to the model's infinite order VAR than in the two variable case. If that conjecture were true we would expect the resulting largest eigenvalue λ_2 in the three variable LSVAR to be much smaller than in the two variable LSVAR. Interestingly, it is not smaller: in the two variable LSVAR it is .97 while in the three variable LSVAR it is .98.

We also experimented with having government consumption be the third shock. The results are very similar. Indeed, Proposition 3 covers this case. (Note that this result is intuitive since the formulas for these eigenvalues do not depend on the parameters governing the added shock.)

In sum, we found that adding the investment-output ratio to our SVARs, as suggested in the literature, does not improve their ability to reproduce the theoretical responses to technology shocks.

8. State Space Approaches as Alternatives

The SVAR methodology holds out the promise of being able to identify technology shocks and their effects with minimal use of economic theory. We have shown that, unfortunately, the specification of SVARs used in practice does not nest most existing business cycle models and that this misspecification leads researchers to draw mistaken inferences. Here we discuss alternative approaches to SVARs to identifying impulse responses to various shocks in the data.

What will clearly not work is to simply add many more lags to the original VAR. Since actual data is not 100,000 observations long, but is less than 200 observations long, it is impossible to accurately estimate lags long enough to get reasonable approximations to the actual impulse response functions.

Another way to proceed might be to expand the list of variables in the VAR to include ones like capital. Unless this approach is guided more specifically by the theories under consideration it is likely to be haphazard. Our conjecture is that this approach will ultimately just bury the most obvious problems but leave more subtle problems intact. We think that a much safer and transparent approach is to use economic theory to guide the specification.

In this vein, the following procedure to specifying models seems the most obvious. First, write down a class of models for which the procedure is supposed to apply. Second, work out the statistical representations implied by the models. Third, figure out some minimal set of identifying restrictions needed to identify the key parameters of the models. In practice, of course, this third step is the difficult one.

As we have shown standard business cycles such as ours have state transition equation

$$(21) \quad S_{t+1} = FS_t + G\varepsilon_{mt+1}$$

and the observation equation

$$(22) \quad X_t = HS_t + u_t.$$

The system (21) and (22) is a standard linear state space system that is easy to estimate by standard statistical methods. In our procedure the parameters of F, G and H all come from the original structural parameters of the model and these matrices have many cross equation restrictions that arise from the underlying economics. One way to proceed which imposes fewer restrictions than the theory would be to apply only the minimum needed to identify the parameters of interest. We could then estimate the system using, say, U.S. data and calculate the impulse responses from the estimated system.

Of course, if we wanted to claim that the impulse responses from such a procedure applied to a broad class of models, including models with sticky prices, models with financial frictions and so on, then we would have to find a state space representation and a set of identifying assumptions that nested the class of models of interest.

An approach that is closely related to the one we have just discussed is the business cycle accounting approach of Chari, Kehoe and McGrattan (2003).

9. Related Literature

In terms of the literature, critiques of the SVAR approach are not new. These critiques can broadly be divided into critiques based on invertibility problems, critiques using economic models as tests, critiques of circular specification searches, and critiques based on deep inference problems when the parameter spaces are infinite-dimensional.

In a pair of insightful but often neglected papers, Hansen and Sargent (1980 and 1991) pointed out that invertibility problems may plague the type of Box-Jenkins methods that underlie the SVAR literature. They showed that interesting economic models could have noninvertible moving average representations and that this noninvertibility could cause problems for simple statistical procedures that do not use enough economic theory. Lippi and Reichlin (1993), along the lines of Hansen and Sargent (1991), analyzed how invertibility problems could lead to mistaken inferences in the Blanchard-Quah procedure. Blanchard and Quah (1993) argue that while such problems may arise for some examples, they typically have not arisen in most applied models. They also argue that even when they do arise the resulting inference problems may not be quantitatively large. As we have argued, our critique is very different from the Hansen–Sargent invertibility critique.

Cooley and Dwyer (1998) lucidly critiqued the SVAR procedure using economic models as tests in a manner broadly similar to ours. One important difference between our work and theirs is that we use estimated models that account well for the U.S. data and they use simpler expository models. Another difference is that we focus on the main conclusion of the recent SVAR literature, namely that technology shocks lead to a fall in hours, while they focus on a variety of other possibly mistaken inferences.

Erceg, Guerrieri, and Gust (2004) also tested the SVAR procedure using economic models. In contrast to our focus on mistaken inferences, their main focus is on small sample bias in SVARs and they conclude that the small sample bias problem in models is modest. Most importantly, they conclude, (p.4), "Overall, Gali's methodology appears to offer a fruitful approach to uncovering the results of technology shocks..." We conclude the opposite.

Uhlig (2004) criticizes the circularity of searching over specifications until a certain pattern is found and then arguing that the data showed that finding such a pattern is strong evidence for a certain theory.

Faust and Leeper (1997) discuss inference problems in infinite-dimensional VARs that underlie the SVAR approach. They argue that (p. 345) "unless strong restrictions are applied, conventional inferences regarding impulse responses will be badly biased in all sample sizes." They show that under a long-run identifying scheme, any test of the magnitude of an impulse response coefficient has a significance level greater than or equal to its power. Faust and Leeper's results build on a pair of seminal papers by Sims (1971, 1972) who showed that in infinite-dimensional spaces, unless severe restrictions are imposed on the parameters, standard methods cannot be used to make asymptotically valid confidence statements.

10. Conclusion

Simple data analysis techniques which reliably point us towards quantitatively promising models can be very useful in applied economic analysis. We have found, however, that the procedure used in the existing technology–focused SVAR literature is not a reliable technique. We have shown that the moving average representations generated by both the differenced and the level SVARs are misspecified with respect to the very business cycle model for which the approach is intended to shed light. Moreover, we how shown that this misspecification leads to quantitatively large mistaken inferences.

Our paper reinforces the point made by Hansen and Sargent (1991) over a decade ago: the main problem with the SVAR approach is that it uses too little a priori economic theory. Without more economic theory it seems to be impossible to determine the answer to basic issues like the ones discussed here: For what questions will a short lag length be reasonable? What variables should we include in the VAR?

Elsewhere, Chari, Kehoe, McGrattan (2003), we have argued that it is both easier and more robust to use an alternative approach called business cycle accounting. This approach has the same goal of the SVAR, namely to quickly shed light on which of a class of models is promising, but it suffers from fewer of the shortcomings. This approach avoids the problems of the SVAR approach by building in enough economic theory to answer the questions it poses.

Another approach, which uses more economic theory than that used in the SVAR procedures but less than the theory used in the business cycle accounting approach, is also likely to be fruitful. Such an approach begins by recognizing that business cycle models have state space representations. This alternative approach involves three steps. The first is to write down a state space representation that nests the class of models of interest. The second is to prove a theorem that a common minimal set of identifying assumptions applies to all models in this class. The third is to estimate the resulting state space model. Either of these alternative approaches seems more likely to lead to progress than the SVAR approach.

References

Anderson, Evan W., Hansen, Lars Peter, McGrattan, Ellen R. and Sargent, Thomas J. 1996. On the mechanics of forming and estimating dynamic linear economies. *Handbook of Computational Economics*, North Holland, Elsevier.

Blanchard, Olivier Jean and Quah, Danny. 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances, *American Economic Review*, v. 79, iss. 4, pp. 655-73.

Blanchard, Olivier Jean and Quah, Danny. 1993. The Dynamic Effects of Aggregate Demand and Supply Disturbances: Reply, *American Economic Review*, v. 83, iss. 3, pp. 653-58.

Chari, V. V., Kehoe, Patrick J., and McGrattan, Ellen. 2004. Business Cycle Accounting, NBER Working Paper 10351.

Christiano, Lawrence J., Eichenbaum, Martin and Vigfusson, Robert. 2003. What Happens After a Technology Shock? NBER Working Paper 9819.

Cooley, Thomas F. and Dwyer, Mark. 1998. Business Cycle Analysis without Much Theory: A Look at Structural VARs, *Journal of Econometrics*, v. 83, iss. 1-2, pp. 57-88.

Cox, D. R. 1961. Tests of Separate Families of Hypotheses, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press: Berkeley, Vol. 1, 105-123.

Erceg, Christopher, Guerrieri, Luca, and Gust, Christopher. 2004. Can long-run restrictions identify technology shocks? Board of Governors of the Federal Reserve System, International Finance Discussion Paper 792.

Faust, Jon. 1996. Near Observational Equivalence and Theoretical Size Problems with Unit Root Tests, *Econometric Theory*, 724-731.

Faust, Jon and Leeper, Eric. 1997. When Do Long-Run Identifying Restrictions Give Reliable Results? Journal of Business and Economic Statistics, vol 15 (3), 345-353.

Francis, Neville and Ramey, Valerie A. 2003. Is the Technology-Driven Real Business

Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited. Mimeo, University of California, San Diego.

Francis, Neville and Ramey, Valerie A. 2004. A New Measure of Hours Per-Capita with Implications for the Technology-Hours Debate. Mimeo, University of California, San Diego.

Fuller, Wayne. 1976. *Introduction to Statistical Time Series*, New York, Wiley.

Gali, Jordi. 1999. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?, *American Economic Review*, v. 89, iss. 1, pp. 249-71.

Gali, Jordi, and Rabanal, Pau. 2004. Technology Shocks and Aggregate Fluctuations: How Well does the RBC Model Fit Postwar U.S. Data? Mimeo, Pompeu Fabra.

Gali, Jordi, Lopez-Salido David J. and Valles, Javier. 2003. Technology Shocks and Monetary Policy: assessing the Fed's performance. Journal of Monetary Economics 50, pp. 723-743.

Hansen, Lars Peter and Sargent, Thomas. 1980. Formulating and Estimating Dynamic Linear Rational Expectations Models, *Journal of Economic Dynamics and Control,* v. 2, iss. 1, pp. 7-46.

Hansen, Lars Peter and Sargent, Thomas J. 1991. Two Difficulties in Interpreting Vector Autoregressions, in *Rational expectations econometrics*, Underground Classics in Economics Boulder and Oxford: Westview Press.

Lippi, Marco and Reichlin, Lucrezia. 1993. The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment, *American Economic Review*, v. 83, iss. 3, pp. 644-52.

McGrattan, Ellen. 1989. Computation and Application of Equilibrium in Models with Distortionary Taxes. PhD thesis Stanford University.

Quah, Danny, 1990. Permanent and Transitory Movements in Labor Income: An Explanation for "Excess Smoothness" in Consumption, *Journal of Political Economy*, v. 98, (3), pp. 449-75.

Shapiro, Matthew D. and Watson, Mark. 1988. Sources of Business Cycle Fluctuations, *NBER Macroeconomics Annual,* pp 111-48.

Sims, Christopher. 1971. Distributed Lag Estimation when the Parameter Space is Explicitly Infinite-Dimensional. *Annals of Mathematical Statistics* 42, 1622-1636.

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

Uhlig, Harald. 2004. What are the effects of monetary policy on output? Results from an agnostic identification procedure. Mimeo, Humbolt University.

Table 1
Parameters of Vector AR(1) Stochastic Process for the Models†

Estimated Using Maximum Likelihood with Data on
Output, Labor, Investment, and Consumption

$$P = \begin{bmatrix} 0 & 0 \\ 0 & \rho \end{bmatrix} \quad Q = \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_\tau \end{bmatrix}, \quad \text{Mean}(s_t) = [\log \bar{z}, \bar{\tau}]$$

	Parameter Estimates				
Source of Hours per Capita	$\log \bar{z}$	$\bar{\tau}$	ρ	σ_z	σ_τ
Prescott and Ueberfeldt (2003)	.00302	.274	.938	.00581	.00764
	(.00032)	(.00536)	(.00727)	(.00032)	(.00035)
Francis and Ramey (2004)	.00293	.301	.937	.00556	.00760
	(.00033)	(.00383)	(.00986)	(.00031)	(.00048)
Christiano et al. (2004)	.00397	.210	.955	.00633	.01011
	(.00037)	(.00408)	(.00440)	(.00031)	(.00036)
Gali and Rabanal (2004)	.00369	.237	.969	.00888	.00983
	(.00042)	(.00469)	(.00280)	(.00035)	(.00039)

† Numbers in parentheses are standard errors.

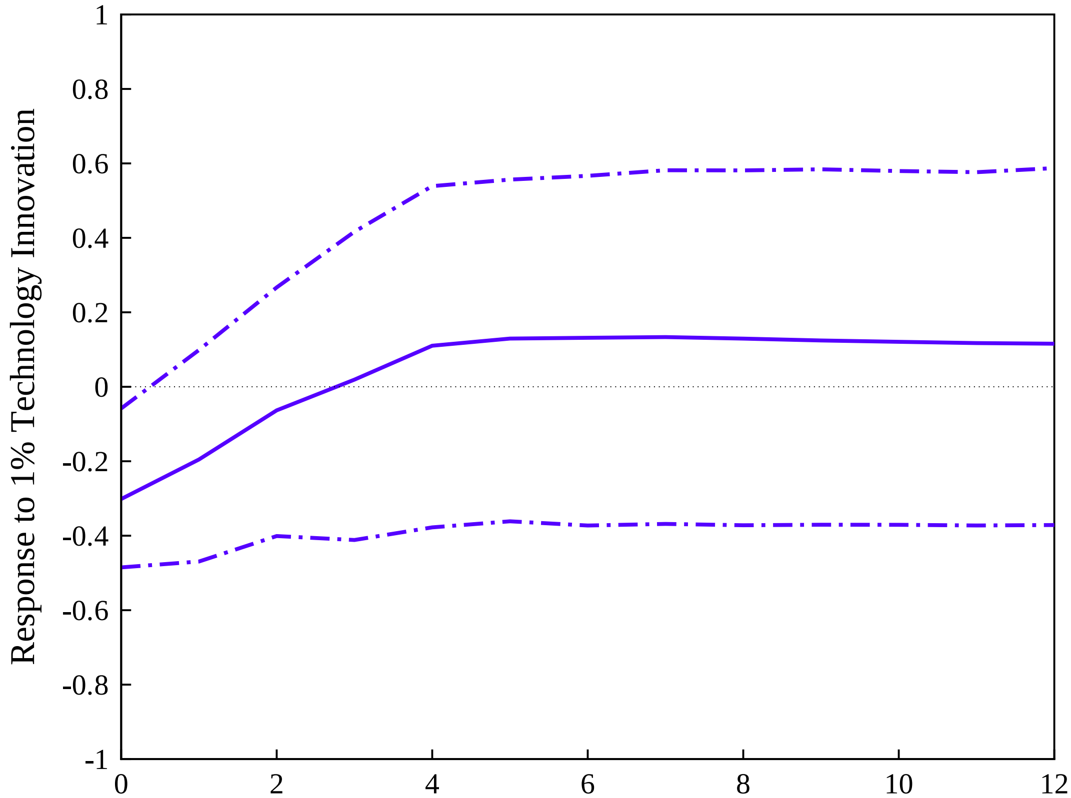

Figure 1
Impulse Response of Hours to Technology Shock Using the DSVAR Procedure with U.S. GDP and Total Hours per capita

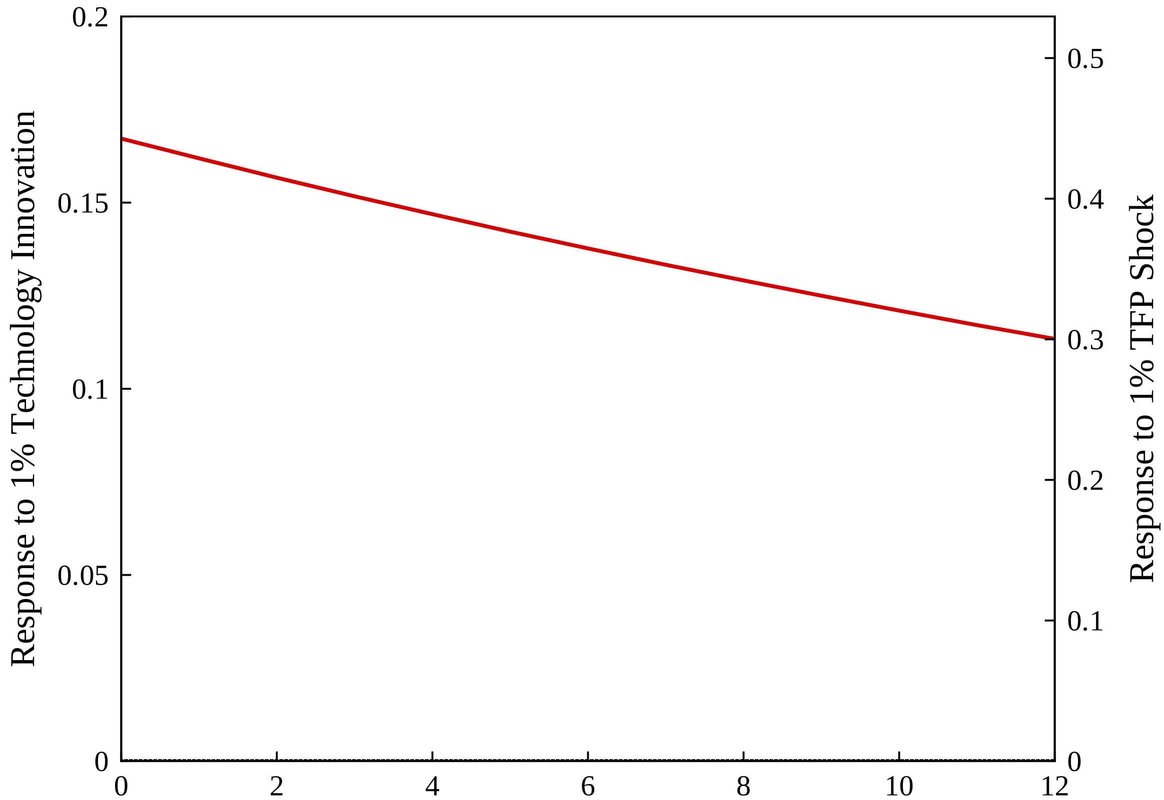

Figure 2
Model Impulse Response of Hours to Technology Shock

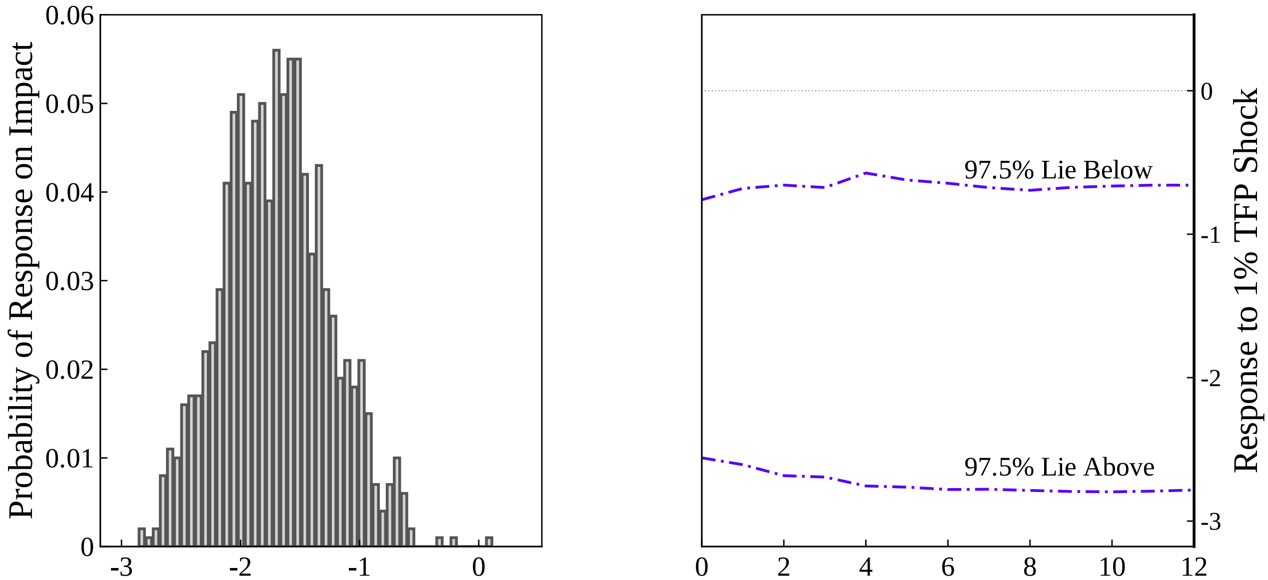

FIGURE 3
Bounds for 95% of Impulse Responses of Hours to Technology Shock
Across 1000 Applications of the DSVAR Procedure to Model Data and
Histogram of Impulse Response in Impact Period

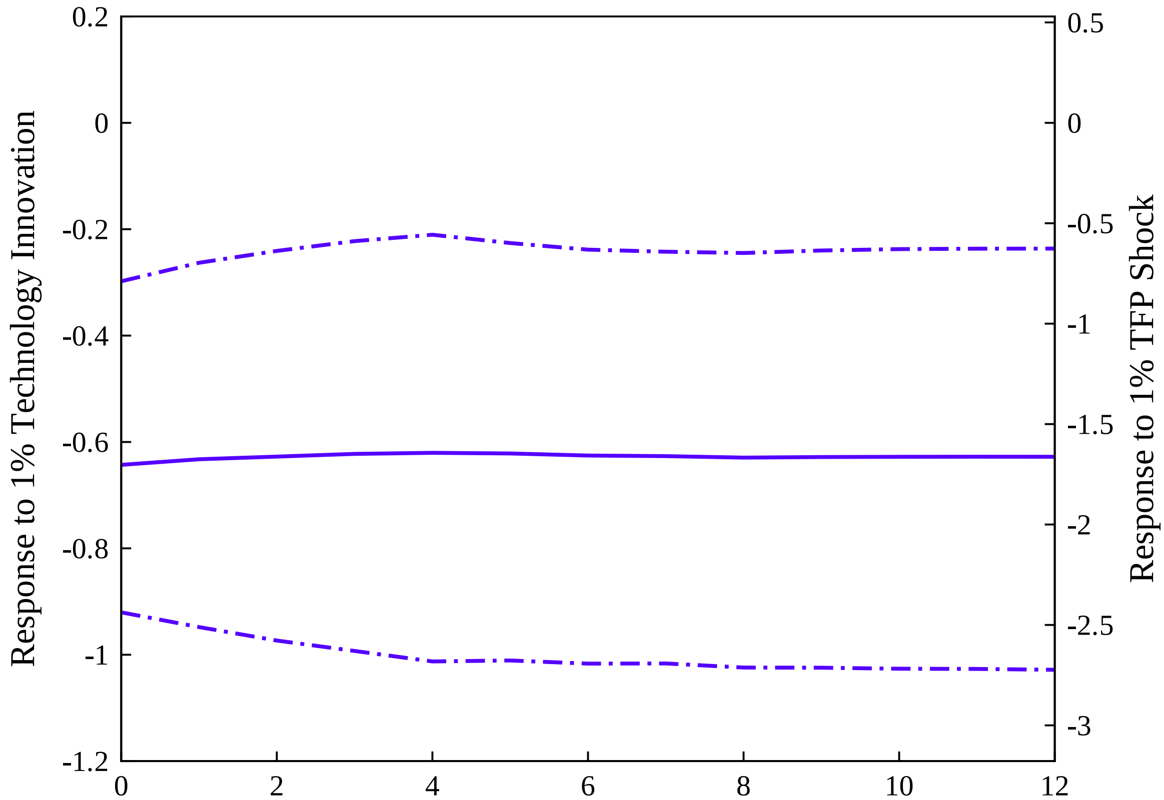

FIGURE 4
Mean Impulse Response of Hours (solid line) and Mean of 95% Bootstrapped Confidence Bands (dashed lines) Averaged Across 1000 Applications of the DSVAR Procedure to Model Data

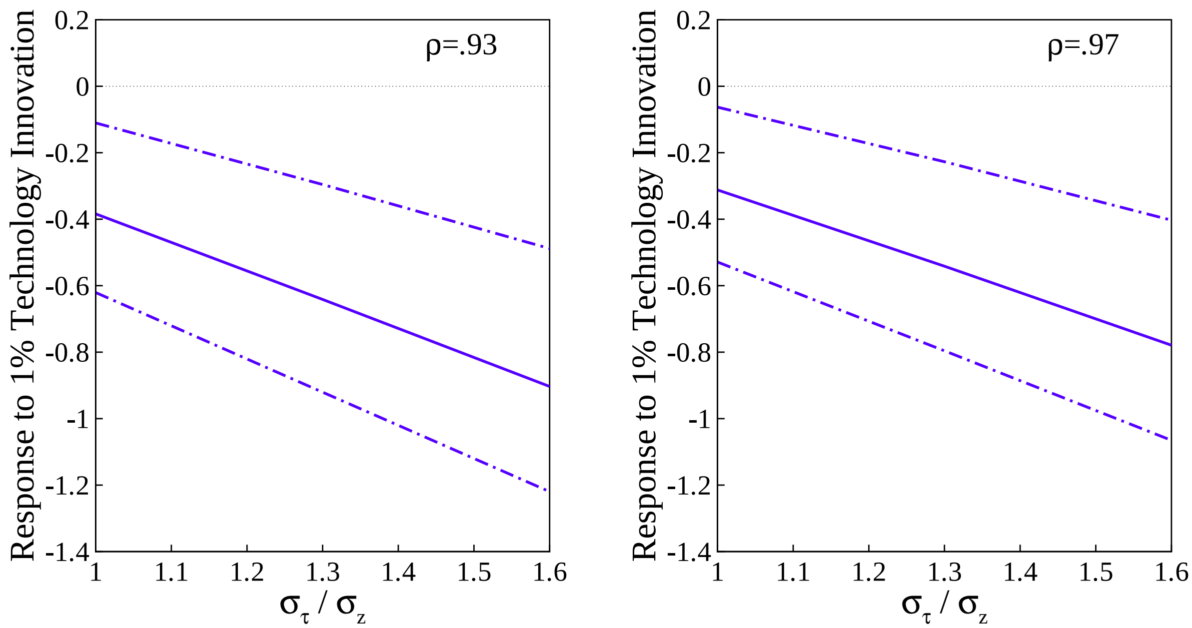

FIGURE 5
Sensitivity of Mean Impulse Response of Hours in Impact Period (solid line) and Mean 95% Bootstrapped Confidence Bands in Impact Period to Variations in ρ and σ_τ/σ_z, with Averages Across 1000 Applications of the DSVAR Procedure to Model Data

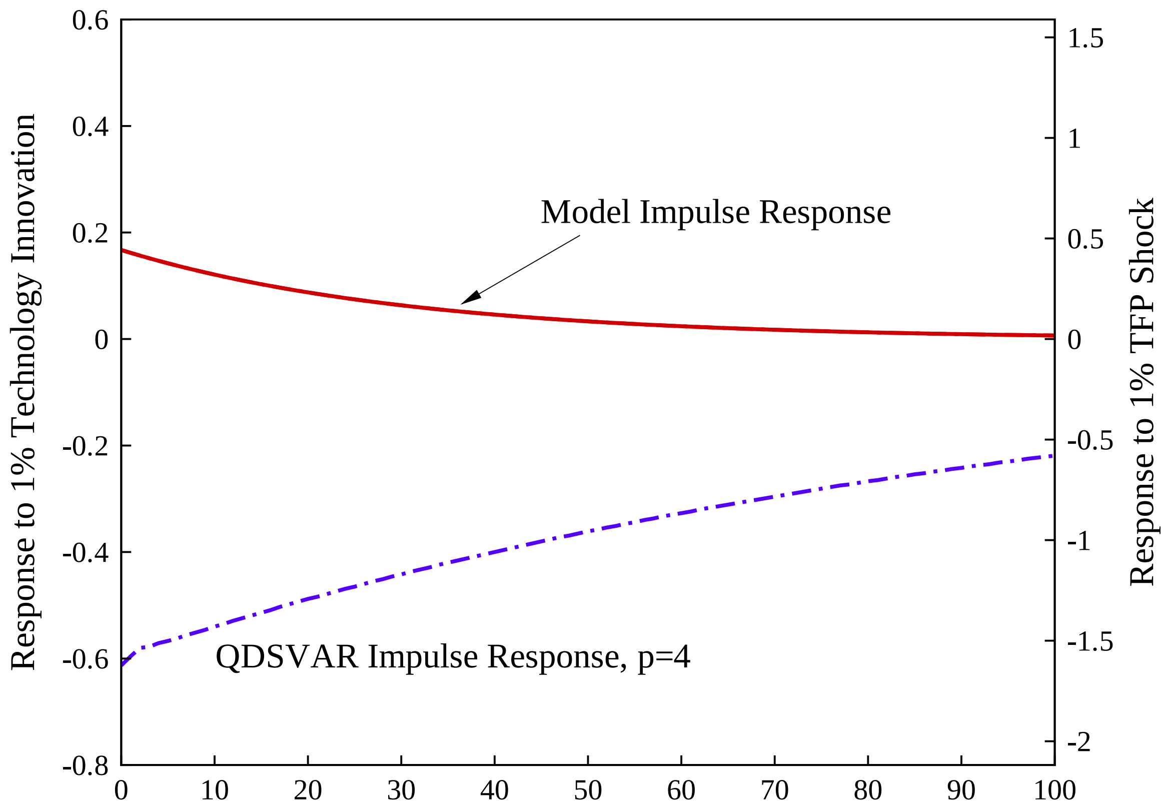

FIGURE 6A
Impulse Responses of Hours for the Model and Those Obtained Using the QDSVAR(.99) Procedure with Four AR Lags Applied to Time Series of Length 100,000 from the Model

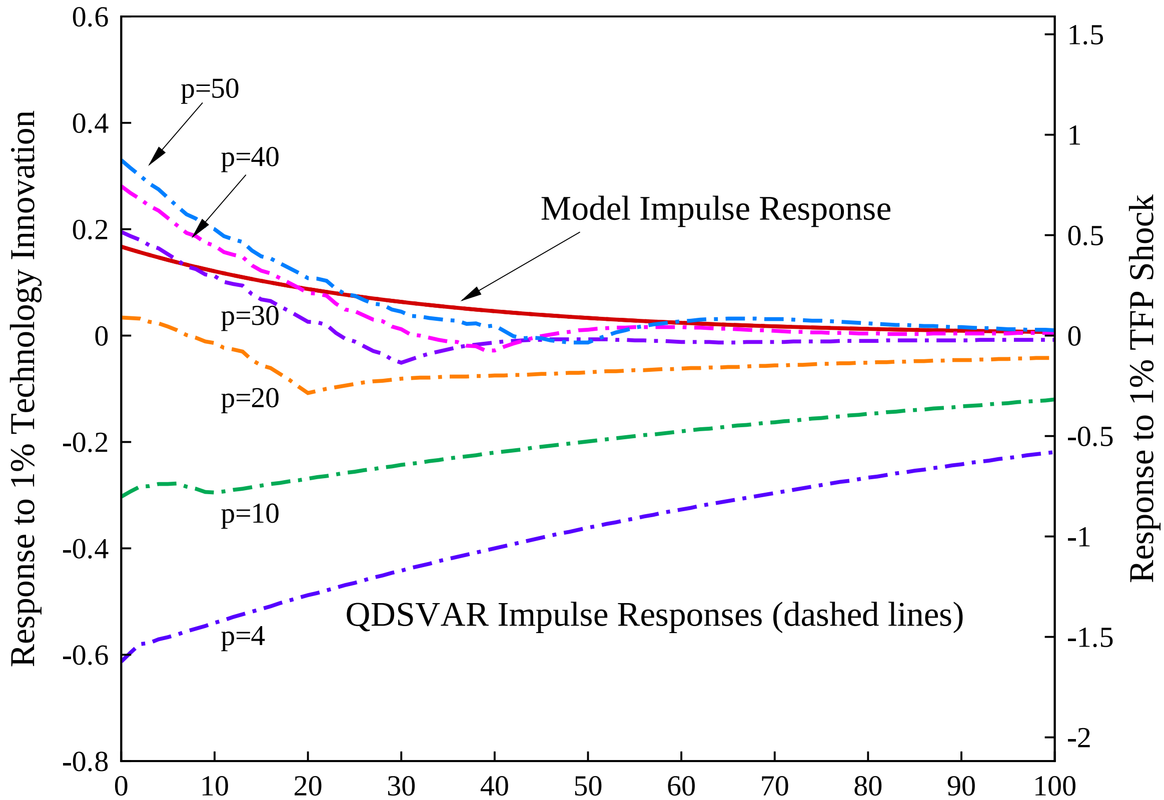

Figure 6b

Impulse Responses of Hours for the Model and Those Obtained Using the QDSVAR(.99) Procedure with Various AR Lags Applied to Time Series of Length 100,000 from the Model

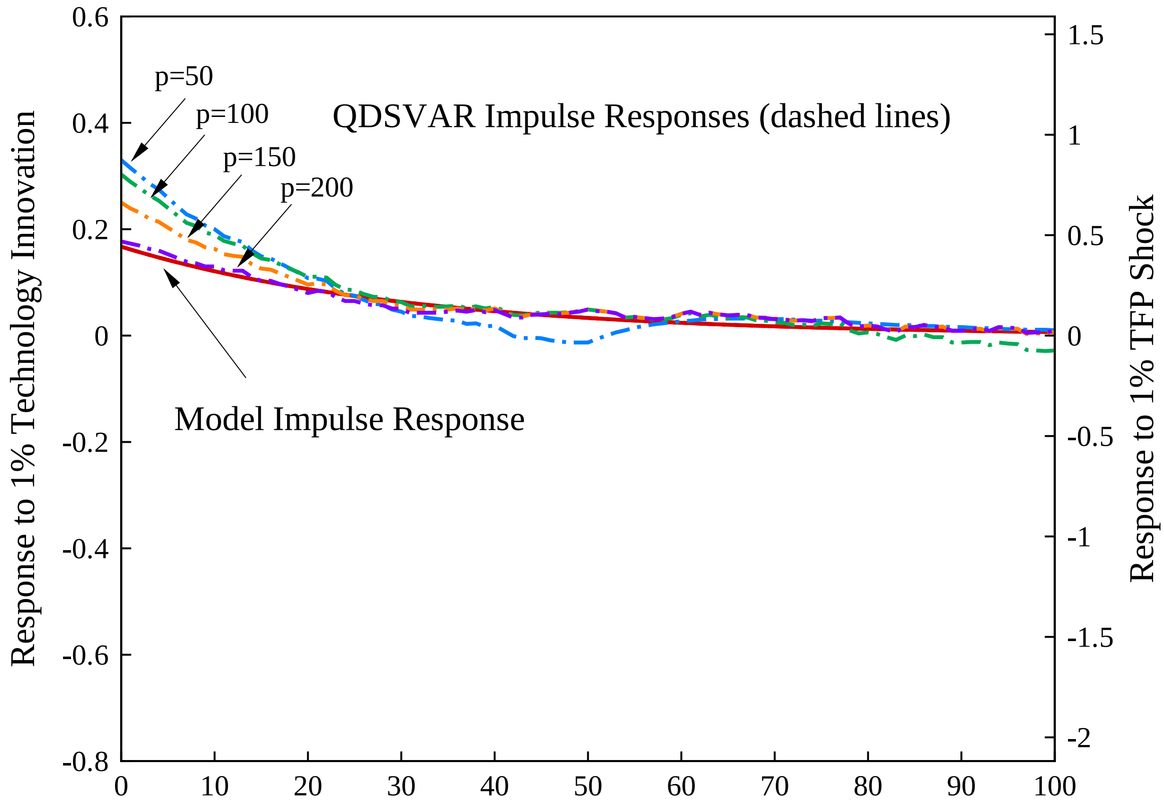

Figure 6c
Impulse Responses of Hours for the Model and Those Obtained Using the QDSVAR(.99) Procedure with Various AR Lags Applied to Time Series of Length 100,000 from the Model

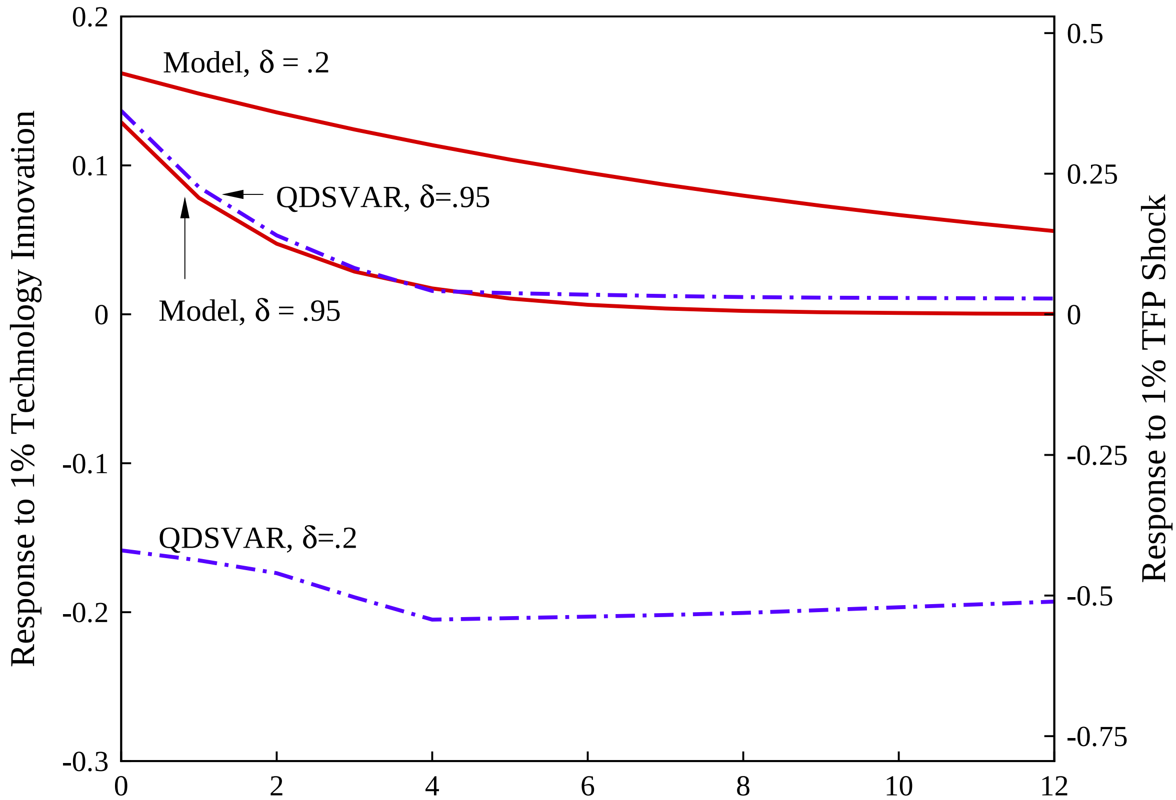

FIGURE 7
Impulse Responses of Hours for the Model with Various Depreciation Rates and Those Obtained Using the QDSVAR(.99) Procedure with Four AR Lags Applied to Time Series of Length 100,000 from the Model

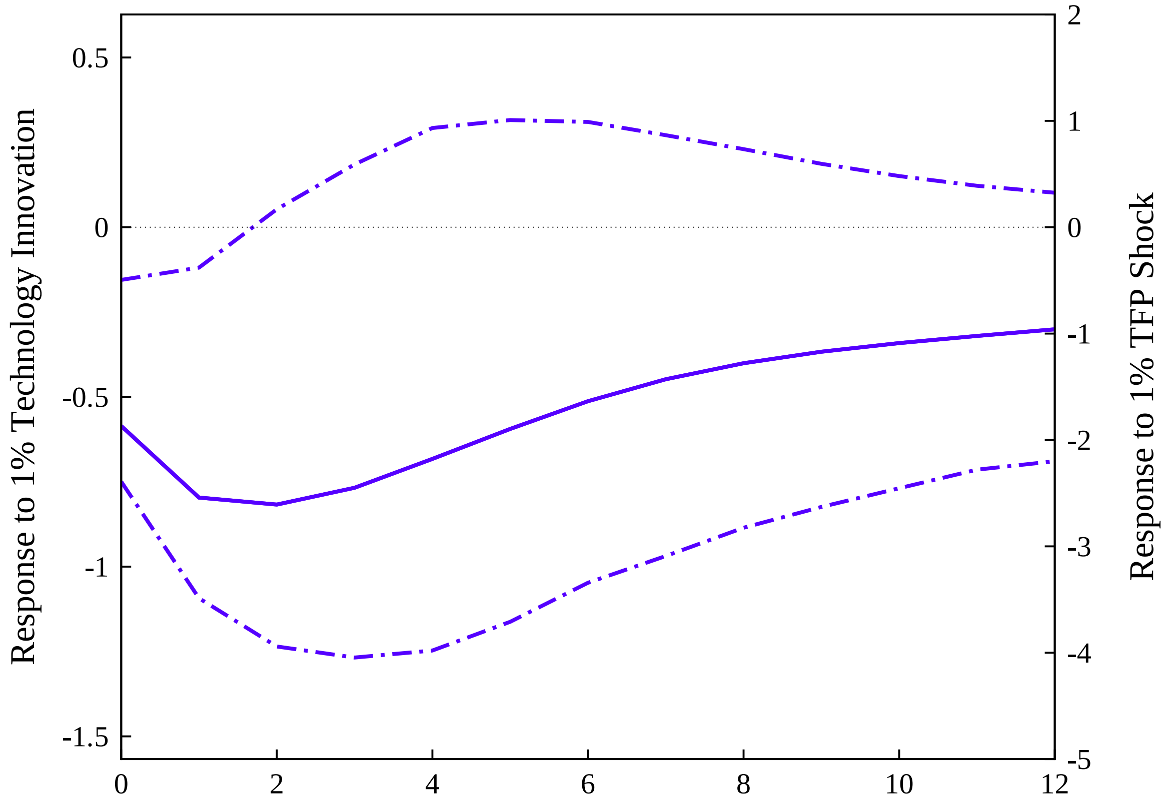

Figure 8a

Impulse Response of Hours to Technology Shock Using the LSVAR Procedure with Dataset of Francis and Ramey

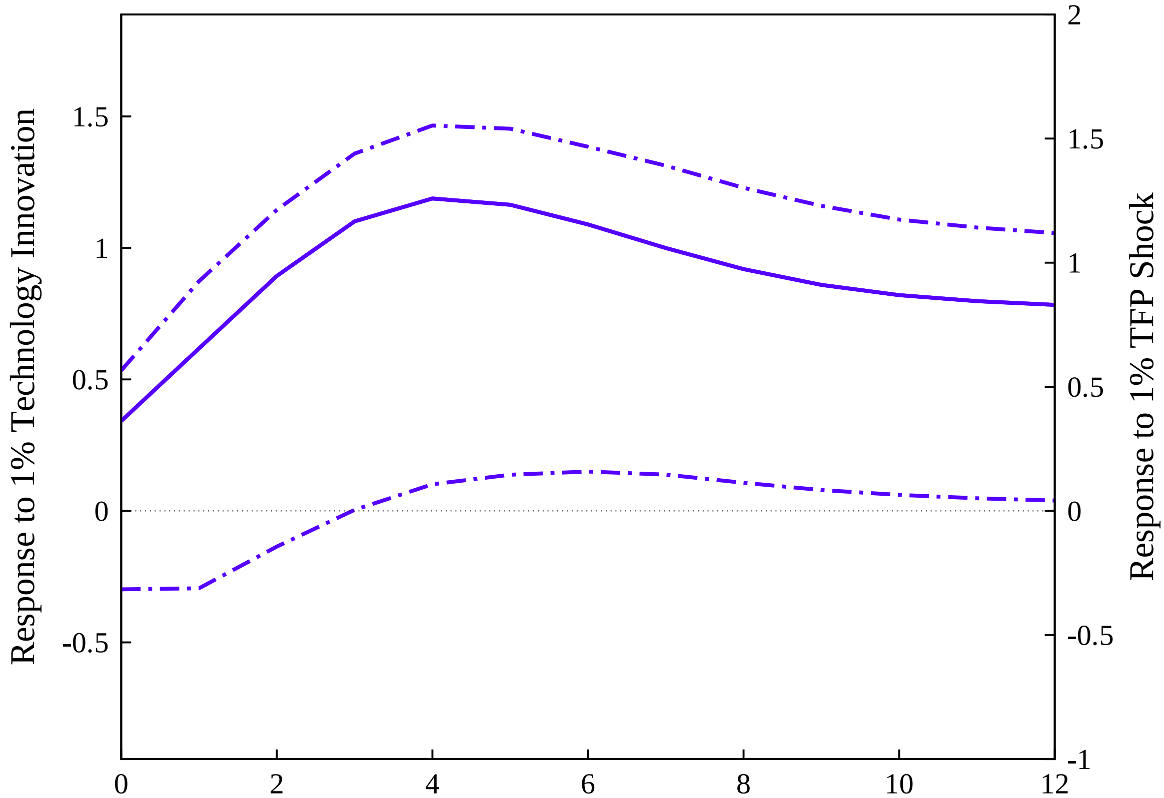

FIGURE 8B
Impulse Response of Hours to Technology Shock Using the LSVAR Procedure with Dataset of Christiano, Eichenbaum, and Vigfusson

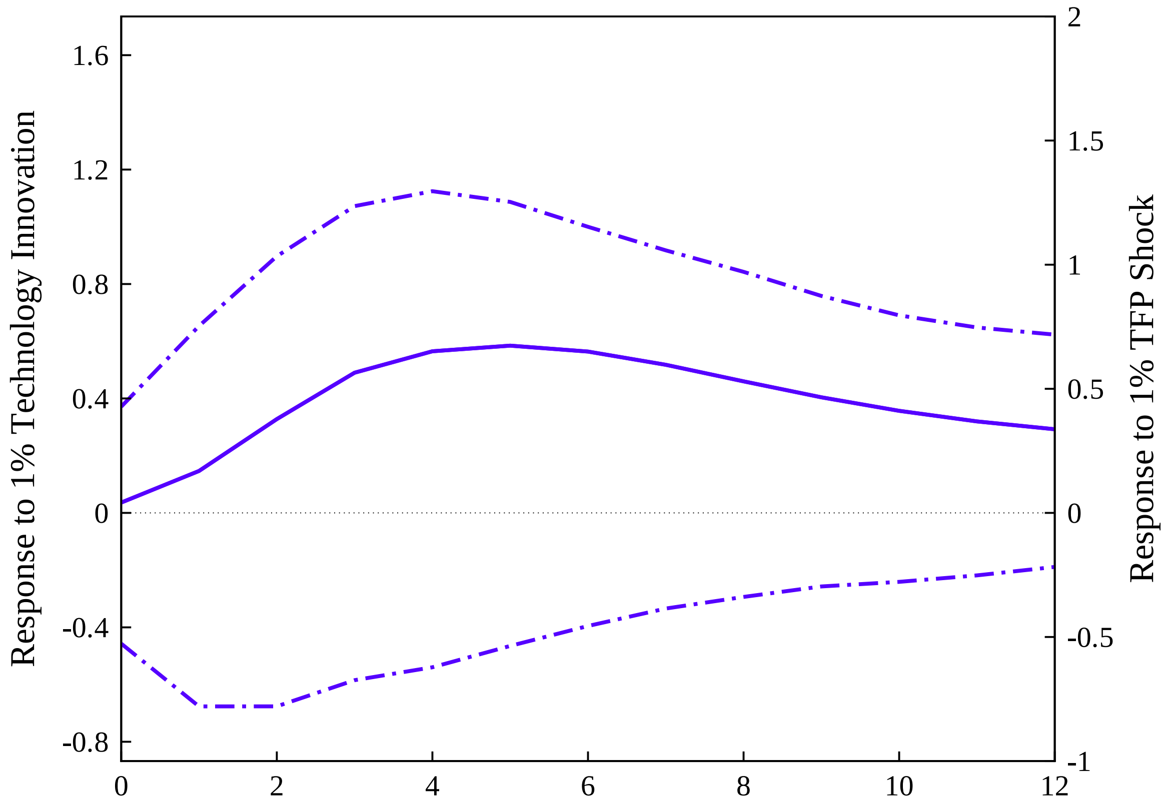

FIGURE 8C
Impulse Response of Hours to Technology Shock Using the LSVAR Procedure with Dataset of Gali and Rabanal

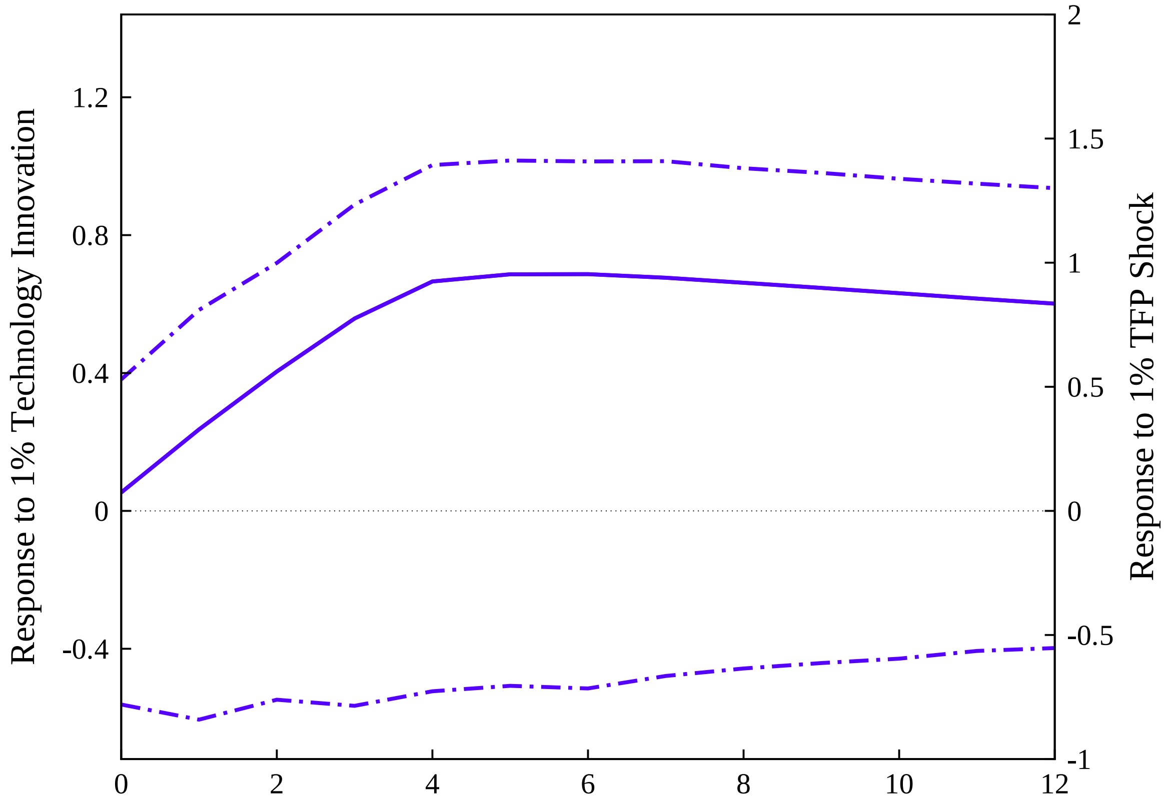

Figure 8d

Impulse Response of Hours to Technology Shock Using the LSVAR Procedure with GDP and Total Hours per capita

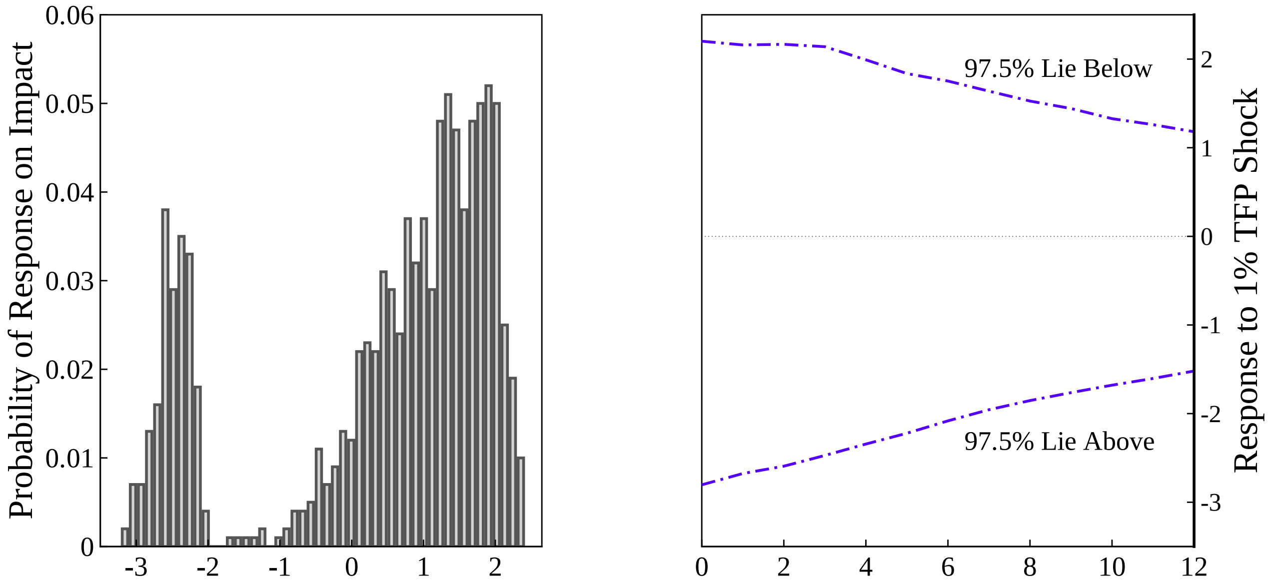

FIGURE 9
Bounds for 95% of Impulse Responses of Hours to Technology Shock
Across 1000 Applications of the LSVAR Procedure to Model Data and
Histogram of Impulse Response in Impact Period

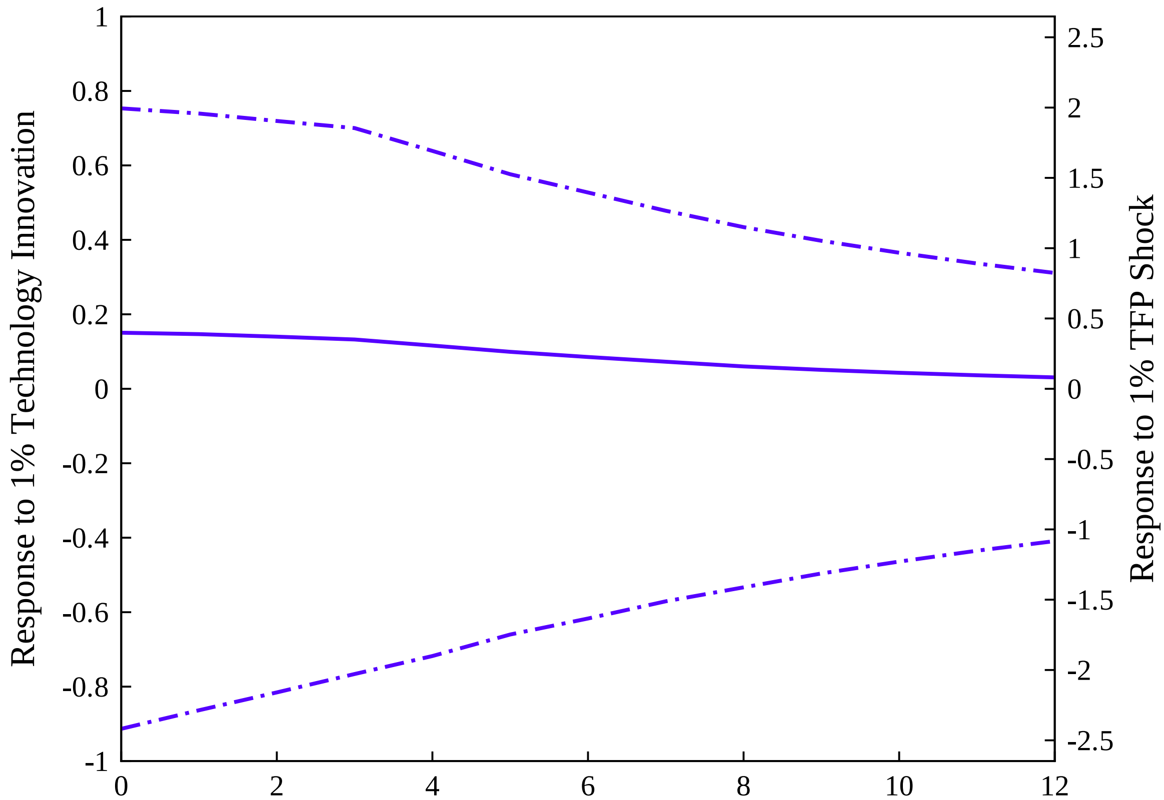

Figure 10
Mean Impulse Response of Hours (solid line) and Mean of 95% Bootstrapped Confidence Bands (dashed lines) Averaged Across 1000 Applications of the LSVAR Procedure to Model Data

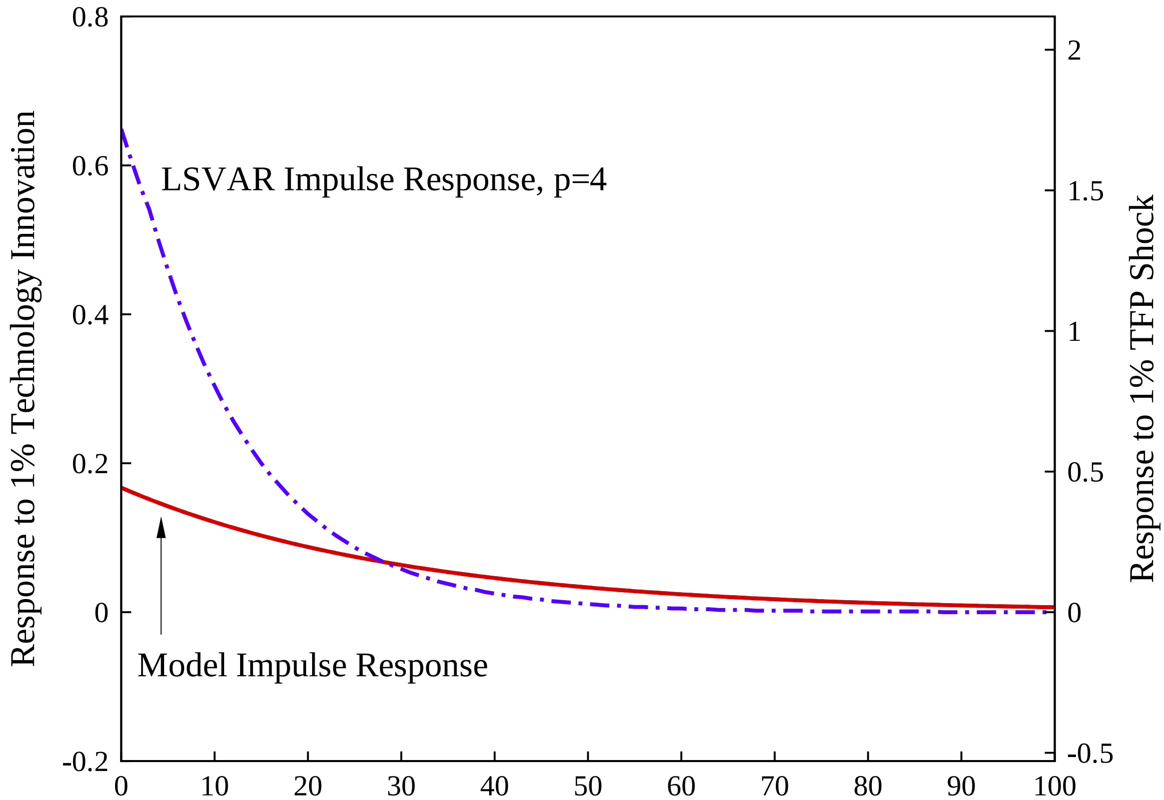

FIGURE 11A
Impulse Responses of Hours for the Model and Those Obtained Using the LSVAR Procedure with Four AR Lags Applied to Time Series of Length 100,000 from the Model

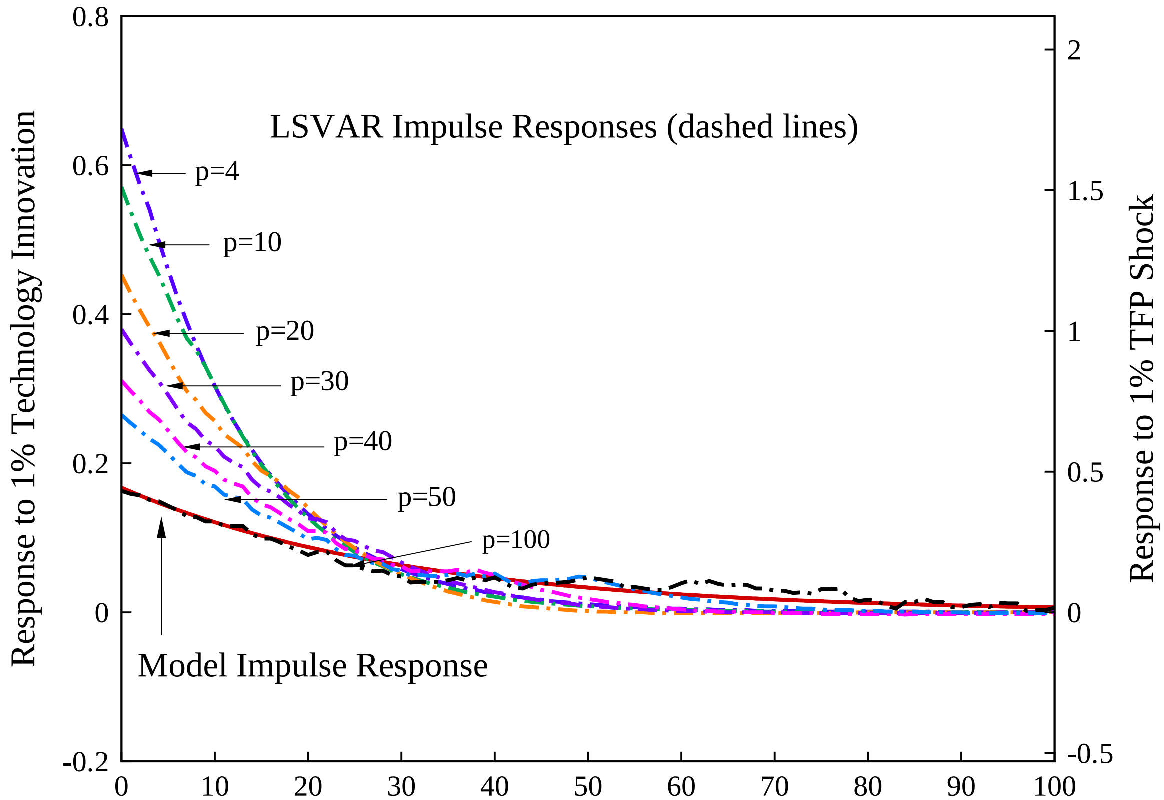

FIGURE 11B

Impulse Responses of Hours for the Model and Those Obtained Using the LSVAR Procedure with Various AR Lags Applied to Time Series of Length 100,000 from the Model

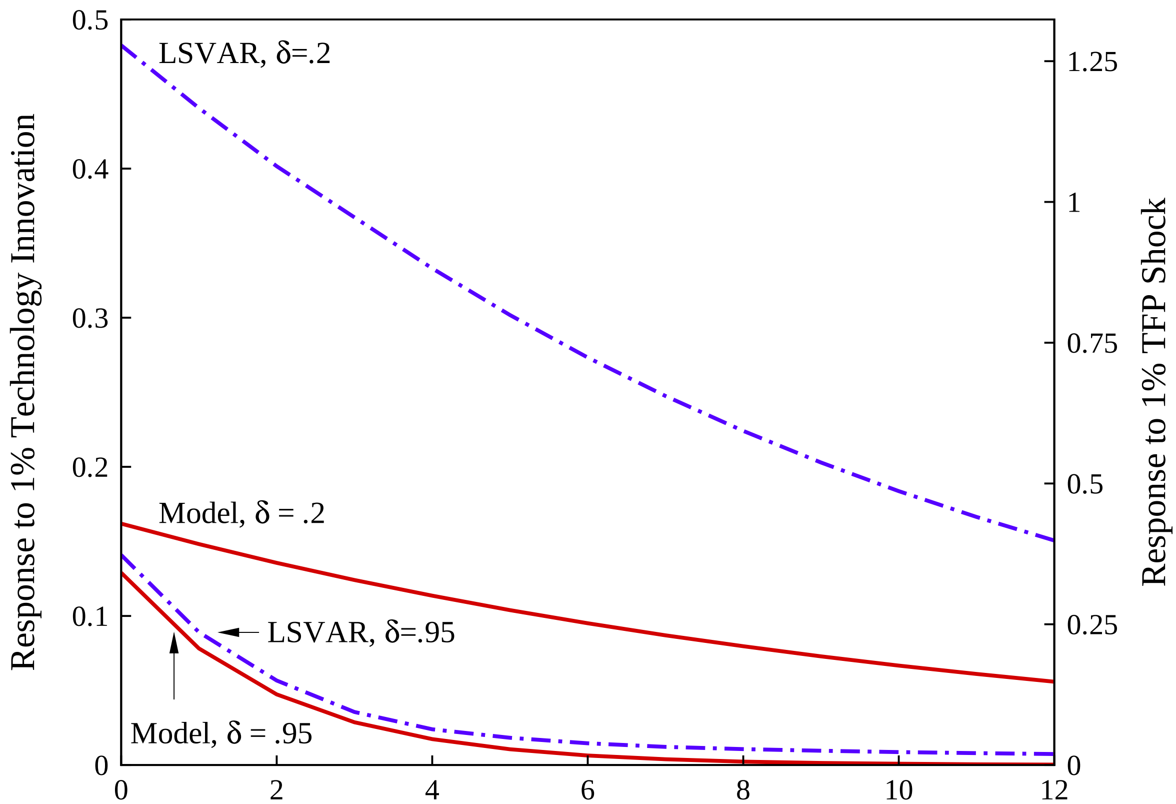

FIGURE 12
Impulse Responses of Hours for the Model with Various Depreciation Rates and Those Obtained Using the LSVAR Procedure with Four AR Lags Applied to Time Series of Length 100,000 from the Model

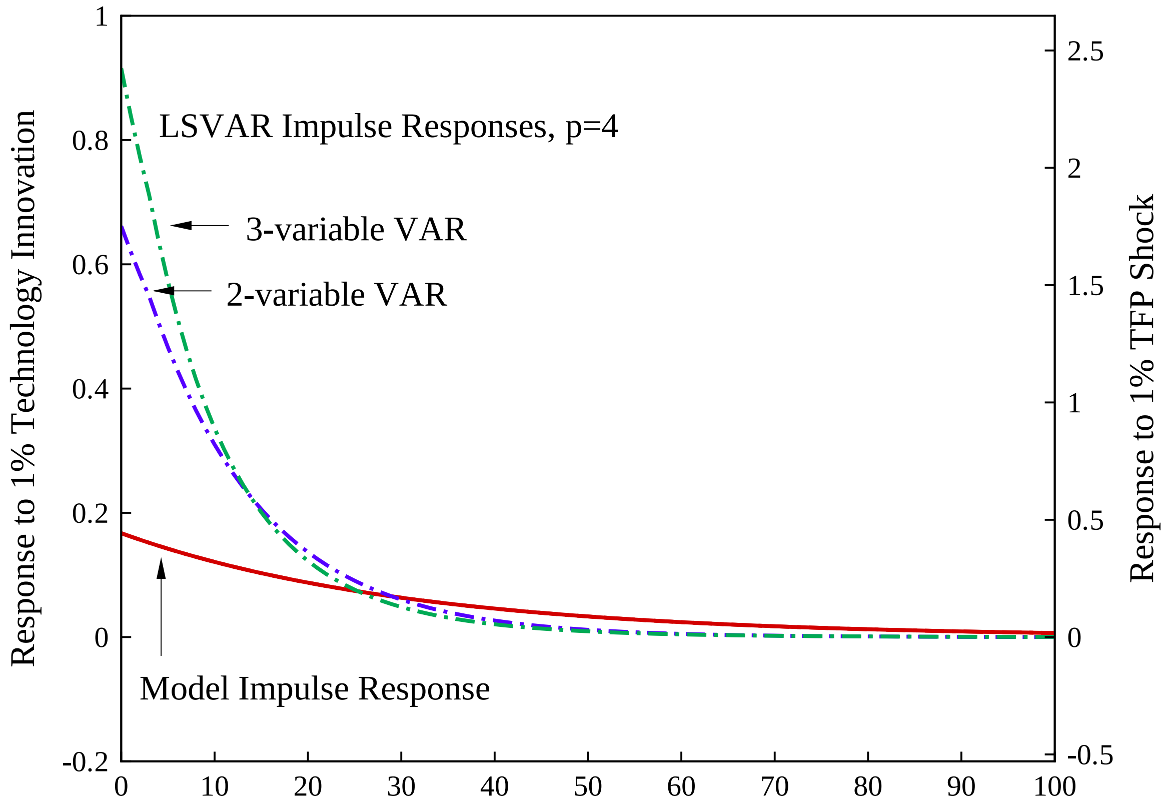

FIGURE 13
Impulse Responses of Hours for the Model and Those Obtained Using the LSVAR Procedure with Four AR Lags Applied to Model Time Series with Either Two or Three Variables in the VAR