Federal Reserve Bank of Minneapolis
Research Department

A Critique of Structural VARs
Using Business Cycle Theory[*]

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

Working Paper 631

Revised May 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 business cycle models in which technology shocks lead to a rise in hours should be discarded. We evaluate the DSVAR approach by asking, is the specification derived from this approach misspecified when the data are generated by the very model the literature is trying to discard? 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 that uses the level of hours (LSVAR) is also misspecified. We argue that alternative state space approaches, including the business cycle accounting approach, are more fruitful techniques for guiding the development of business cycle theory.

The goal of the structural vector autoregression (SVAR) approach is to identify promising classes of business cycle models using a simple time series procedure. The idea behind the procedure is to run vector autoregressions in the data and impose identifying assumptions to back out impulse responses to various shocks.[1] These SVAR impulse responses are then, typically implicitly, compared with theoretical impulse responses from economic models. Importantly, this literature does not follow the procedure that Sims (1989) advocates, in which the same VAR is run on data from an actual economy as on data generated from a model and statistics from the VARs are compared.

We focus on the branch of the literature that studies impulse responses to a technology shock. We subject the SVAR procedure to a natural economic test. We treat an economic model as the data-generating mechanism and calculate the population impulse responses obtained from applying the SVAR procedure to data from the model. We ask whether the impulse responses identified by the SVAR procedure are close to the model's impulse responses. For a large class of parameters, including ones estimated from the data, we find they are not. In this sense, we provide counterexamples to claims in the literature that, as long as the model satisfies the key identifying assumptions, the procedure will uncover the model's impulse responses. In addition, we show analytically when the SVAR procedure will produce responses close to the model responses and when it will not.

This technology shock branch of the SVAR literature 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, and the level specification, called the *LSVAR*, uses the level of hours. Both branches of the SVAR literature make several assumptions to identify the underlying shocks, often labeled as *demand shocks* and *technology shocks*. This literature views two identifying assumptions as key: (*i*) demand shocks have no permanent

---

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

effect on the level of labor productivity while technology shocks do, and (*ii*) the demand and technology shocks are orthogonal. Both branches estimate a VAR with a small number of lags, typically four.

The main finding of the DSVAR literature is that a technology shock leads to a fall in hours. Gali (1999), Francis and Ramey (2003), and Gali and Rabanal (2004) use the DSVAR procedure to infer that this finding 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.

In the LSVAR literature researchers report a wide range of results. Francis and Ramey (2004) argue that the LSVAR evidence shows that real business cycle models are dead. Conversely, Christiano, Eichenbaum, and Vigfusson (2003) maintain that their LSVAR results imply that these models are alive and well, while Gali and Rabanal (2004) assert that their LSVAR results, by themselves, are inconclusive. As we document below, these sharply contrasting results are driven almost entirely by small differences in the underlying data.

We tilt our test in favor of the SVAR procedure by focusing mainly on a stripped-down business cycle model, referred to as the *baseline model*, that satisfies the two key identifying assumptions of the SVAR literature. Most business cycle models do not satisfy these two assumptions. In this sense our model is a best-case scenario for the SVAR procedure. Our estimated model shares the feature of second generation business cycle models in that it has multiple shocks with stochastic processes estimated from the data.

In order to abstract from small sample biases and sampling uncertainty, we mainly study population impulse responses obtained from applying the SVAR procedure to our model rather than impulse responses computed from short samples generated by our model. The population impulse responses from the DSVAR procedure imply that a technology shock leads to a decline in hours. This conclusion is mistaken because in our model a technology shock leads to a rise in hours. The population impulse responses from the LSVAR procedure imply that a technology shock leads to a rise in hours about three times that in the model.

2

In this sense the LSVAR procedure also leads to a mistaken conclusion.

We deconstruct the mistaken conclusion and find that the source is a misspecification of the SVAR. This misspecification arises not because the model fails to satisfy the key identifying assumptions but because the model's VAR fails an auxiliary assumption. This auxiliary assumption is that the stochastic process for labor productivity and hours is well approximated by an autoregressive representation with a small number of lags.

Our deconstruction rests on four propositions. In our propositions we nest the DSVAR and LSVAR specifications in a specification called the *nested* SVAR, in which hours, l_t, are quasi-differenced (with $\log l_t - \alpha \log l_{t-1}$). The first proposition shows that the VAR of a standard business cycle model in labor productivity and quasi-differenced hours is of infinite order and the VAR coefficients decay very slowly. More precisely, after the first lag, the VAR coefficient matrices at any lag are equal to the product of a fixed *decay matrix* and the coefficient matrix of the previous lag. The largest eigenvalue of the decay matrix is close to one.

In the second proposition we derive key formulas that determine the *technology shock specification error*, defined as the difference between the population impulse response for the nested SVAR and the model's impulse response to a technology shock. We show that for a range of parameter values, including our estimated parameter values, this specification error is large even when the VAR has over 50 lags.

The third proposition shows that the SVAR procedure uncovers the model's impulse response in two cases: the economy has no capital or the economy has only one shock. When the economy has no capital, the coefficients of the model's VAR associated with technology shocks are zero and hence are well modeled with a small number of lags. When the economy has only one shock, the effects of various shocks cannot be confounded with each other and the SVAR procedure also uncovers the model's impulse response. More generally, we show that the technology shock specification error is smaller the less important is capital and the

3

less important is the demand shock.

Our finding for the link between the technology shock specification error and the size of the demand shock implies that the SVAR procedure performs relatively well for first generation models of the business cycle in which technology shocks account for the bulk of fluctuations in output and that it performs relatively poorly for second generation models that emphasize the role of other shocks in accounting for fluctuations in a broader set of aggregates.

We investigate whether adding other variables to the SVAR, such as the investment-output ratio or the consumption-output ratio, changes our findings. One motivation for adding variables to the SVAR is that in the SVAR literature, researchers often investigate the effects of doing so. When we add other variables to the SVAR, we add other shocks to the model because the covariance matrix of the observed variables is not singular. When we add shocks to the model with stochastic processes estimated from the data, the technology shock specification error remains large. Some intuition for our findings comes from our fourth proposition, in which we show that the largest eigenvalue of the decay matrix for the three- and four-variable systems is still close to one.

We show that under extreme circumstances, in which the variances of the added shocks are much smaller than we find them to be in the data, the technology shock specification error is small. In higher-dimensional versions of these extreme circumstances, we find that the SVAR procedure uncovers the model's impulse response when our *singularity rule* holds, namely, the sum of the number of singularities in the decay matrix of the VAR coefficients and the shock covariance matrix is at least as large as the number of variables in the system.

In order to address small sample bias, sampling uncertainty, and the ability of standard tests to detect the need for long lags, we also examine the SVAR procedure in samples as long as in quarterly, postwar U.S. data. We decompose the *overall difference* between the model's impulse response and the mean across short samples of the SVAR impulse responses into

two parts: the specification error (with four lags), and the *small sample bias,* the difference between the population SVAR impulse responses and the mean across short samples of the SVAR impulse responses.

Consider first the DSVAR specification. With this specification it turns out that the model's moving average representation is at the edge of the noninvertibility region. To overcome this technical problem, we consider a specification that is close to the DSVAR specification in which the quasi-differencing parameter is close to one so that the moving average representation is invertible. We call this specification the *QDSVAR specification.* With it we find that the small sample bias is small compared to the specification error, so the overall difference between the model's impulse response and the mean across short samples of the SVAR impulse response is large. Here the procedure leads to a mistaken inference in that standard tests reject that the impulse responses from the QDSVAR coincide with the model's impulse responses. We also find that standard lag length tests do not detect the need for long lags.

Consider next the LSVAR specification. With this specification, the small sample bias is sizeable and of the opposite sign as the specification error, but the overall difference is still large. For example, for the impact coefficient, the mean of the short sample impulse is more than twice as large as that of the model. In terms of the size of sampling uncertainty, the range of impulse responses across simulations is wide. The associated confidence bands are so wide that the procedure cannot distinguish between large classes of models, including sticky price models and real business cycle models. We also find that standard lag length tests do not detect the need for long lags.

We have shown that the SVAR procedure, as implemented in the literature, is misspecified with respect to the very models on which the procedure is meant to shed light. Of course, statistical procedures are often technically misspecified. Our contribution is to show that this misspecification is quantitatively large for a range of parameters for a business cycle

model that meets the key identifying assumptions of the SVAR procedure.

In this vein, one critique of the DSVAR 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 to 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.[2]

We are sympathetic to the basic vision of the SVAR approach: to develop simple data analysis techniques that apply to as broad a class of models as possible and to help select promising models from that class. We argue that state space methods are relatively simple techniques that apply to broad classes of models but do not suffer from the deficiencies of the SVAR procedures.

We are also sympathetic to applications of VARs using the approach advocated by Sims (1989) and applied by Cogley and Nason (1995), among others. In that approach, the same VAR is run on the data generated from the model and the data from an actual economy, and statistics from the two are compared. The crux of the problem with the SVAR approach is that the results from a regression on the data are not compared with the analogous regression in the model but are instead compared with theoretical impulse responses. As we show, the differences between the analogous regression in the model and the theoretical impulse responses are often quantitatively large.

In our test we use data generated from an economic model because in the model we can

---

[2]Moreover, Marcet (2004) proves that even if the data are stationary, estimating a VAR in differences yields consistent estimates and that standard asymptotic inference is valid. He goes on to argue that in many cases specifying the VAR in differences is a better alternative than specifying it in levels, even when the original data are stationary.

take a clear stand on what constitutes a technology shock. Hence, the question of whether fluctuations in total 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 productivity in the U.S. data fluctuates.

Our critique builds on those in a number of papers that we discuss below, especially Sims (1971, 1972), 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), Francis and Ramey (2003), and Gali and Rabanal (2004). The procedure starts with a vector autoregression of the form

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

where the error terms have variance-covariance matrix $E v_t v_t' = \Omega$ and are orthogonal at all leads and lags, so that $E v_t v_s' = 0$ 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 $x_{2t} = \log l_t - \alpha \log l_{t-1}$, where l_t is a measure of the labor input. We consider two different specifications used in the literature: in the differenced specification (DSVAR) α equals 1, so x_{2t} is the first difference in the log of the labor input; in the level specification (LSVAR) α equals 0, so 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. Inverting this vector autoregression is convenient in order to express it in its equivalent moving average

form,

$$(2) \quad X_t = C_0 v_t + C_1 v_{t-1} + C_2 v_{t-2} + \dots,$$

where $C_0 = I, C_1 = B_1, C_2 = B_1C_1 + B_2, C_3 = B_1C_2 + B_2C_1 + B_3$, 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 structural model is given by

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

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 response of X_t in period $t+i$ to a shock in period t is given by A_i. From these responses the impulse responses for y_t/l_t and l_t can be determined. Note that the impulse responses to a technology shock depend only on the first column of the matrices A_i for $i = 0$, $1, \dots$.

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 $A_i = C_i A_0$ for $i \geq 1$. 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 two 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 $E\varepsilon_t\varepsilon_t' = I$, or equivalently as

$A_0 A_0' = \Omega$. 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 captures 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 assumptions determine the shocks up to a sign convention. Our sign convention is that a technology shock is called *positive* if it raises the level of labor productivity in the long run. The impulse responses for a technology shock are invariant to the sign convention with respect to the demand shock. Since we focus exclusively on the impulse responses for a technology shock, the sign convention for the demand shock is irrelevant for our results.

In the literature, the impulse responses are computed as follows. The problem of determining the matrix A_0 that satisfies the long-run restriction (4), the sign convention on the technology shock, and the orthogonality assumption $A_0 A_0' = \Omega$ is compactly written as follows. Let $\bar{C} = \sum_{i=0}^{\infty} C_i$ be the sum of the moving average coefficients and note that since $A_i = C_i A_0$, $\sum_{i=0}^{\infty} A_i = \bar{C} A_0$. The long-run restriction is that the $(1,2)$ element of the matrix $\bar{C} A_0$ is zero, and the sign convention is that the $(1,1)$ element of $\bar{C} A_0$ is positive. The orthogonality assumption is equivalent to $(\bar{C} A_0)(\bar{C} A_0)' = \bar{C} \Omega \bar{C}'$. The problem reduces to finding A_0 to solve

$$(5) \quad \bar{C} A_0 = L,$$

where L is a lower triangular matrix with a positive $(1,1)$ element that satisfies $LL' = \bar{C} \Omega \bar{C}'$. This procedure uniquely pins down the first column of A_0. Since $A_i = C_i A_0$ this procedure also uniquely pins down the first column of A_i and hence the impulse response to a technology

shock. (This procedure does not uniquely pin down the second column of A_0. To do so we need to impose a sign convention on the demand shock.)

2. An Estimated Multiple Shock Business Cycle Model

Our test uses versions of a business cycle model we think of as second generation because the model has multiple shocks with stochastic processes estimated from the data. (See, among a host of others, McGrattan (1994).) 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. 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 $\bar{g} = g_t/Z_t$. Below we discuss variants of this model in which the investment tax and government spending are stochastic.

Our choice of the labor tax as the second shock is motivated by an extensive literature on second generation business cycle models. This literature grew out of the first generation literature on equilibrium business cycle models. The first generation literature focused on models in which technology shocks accounted for most of the fluctuations in output. (See, for example, Kydland and Prescott (1982) and Hansen (1985).) The second generation models introduced other shocks and were motivated, in part, by the inability of the first generation models to generate the volatility of hours observed in the data.[3]

A key feature of the shocks that many of these models introduced is that they effectively distort consumers' labor-leisure choice. Chari, Kehoe, and McGrattan (2004) show that many of these models, including sticky price models, are equivalent to a prototype business

---

[3]See, for example, Cooley and Hansen (1989), Benhabib, Rogerson, and Wright (1991), Greenwood and Hercowitz (1991), Bencivenga (1992), Rotemberg and Woodford (1992), Braun (1994), McGrattan (1994), Stockman and Tesar (1995), Hall (1997), Bernanke, Gertler, and Gilchrist (1999), and Christiano, Eichenbaum, and Evans (2005.)

cycle model with a labor wedge that resembles a stochastic tax on labor. They also show that the labor wedge and the productivity shock account for the bulk of fluctuations in U.S. data. These considerations lead us to use the labor tax as the second shock. It is worth noting that, in second generation models technology shocks typically account for a much smaller fraction of output variability than they do in first generation models. Our estimated models described below also have this feature.

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

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

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) $\quad \log Z_{t+1} = \mu_z + \log Z_t + \sigma_z \varepsilon_{mt+1}^z$

(9) $\quad \tau_{lt+1} = (1 - \rho_l)\bar{\tau}_l + \rho_l \tau_{lt} + \sigma_l \varepsilon_{mt+1}^d,$

where ε_{mt}^z and ε_{mt}^d 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 $\sigma_z \varepsilon_{mt}^z$ and $\sigma_l \varepsilon_{mt}^d$ 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 parameter ρ_l is the persistence parameter for the labor tax, and the mean of the labor tax is $\bar{\tau}_l$.

Our model meets the two key identifying assumptions of the technology shock SVAR literature. In the steady state of our model, labor productivity is not affected by labor tax rates but is affected by the level of technology. Thus, a shock to labor taxes has no effect on labor productivity in the long run, regardless of the persistence of the stochastic process on labor taxes. Also, by construction, the two shocks are orthogonal.

We use functional forms and parameter values familiar from the business cycle literature for our quarterly model. We assume that the production function has the form $F(k, l) = k^\theta l^{1-\theta}$ and that the utility function has the form $U(c, l) = \log c + \phi \log(1 - l)$. We choose the capital share $\theta = .33$ and the time allocation parameter $\phi = 2.5$. We choose the depreciation rate, the discount factor, and growth rates so that, on an annualized basis, depreciation is 6%, the rate of time preference is 2%, and the population growth rate is 1%.

In estimating the stochastic processes for our model, we use no direct data on tax rates. We follow Chari, Kehoe, and McGrattan (2004) in treating tax rates on labor and investment as unobserved and inferring them from the model's equilibrium conditions. 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 $S_t = (\log \hat{k}_t, s_t, \log \hat{k}_{t-1}, s_{t-1}, 1)$. The state's evolution is determined by the capital stock decision rule,

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

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

$$(11) \quad s_{t+1} = \bar{P} + P s_t + Q \varepsilon_{mt+1},$$

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

$$(12) \quad S_{t+1} = F S_t + G \varepsilon_{mt+1}.$$

Let the observable variables $Y_t = (\Delta \log y_t, \log l_t, \Delta \log x_t, \Delta \log g_t)'$ denote the growth rate of output, the log level of labor, the growth rate of investment, and the growth rate of government spending.[4] The decision rule for labor in the model has the form

$$(13) \quad \log l_t = \psi_0 + \psi_k \log \hat{k}_t + \psi_z \log z_t + \psi_l \tau_{lt},$$

and the decision rule for the growth rate of output has a similar form. We can write the observed variables $Y_t = H S_t$, 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 but four observed variables, we add a small amount of measurement error u_t to what is observed and write the observer equation as

$$(14) \quad Y_t = H S_t + u_t.$$

Specifically, we choose the measurement error of each variable in Y_t to be independent over

---

[4] We measure government spending as the sum of government consumption and net exports. See Chari, Kehoe, and McGrattan (2005) for the rationale for this measure of government spending.

time and across variables with variance one-hundredth of the variance of the corresponding variables in Y_t.

We then use the maximum likelihood procedure described in McGrattan (1994) and in Anderson et al. (1996) to estimate the parameters of the vector AR1 process, (11), using data on output, hours, investment, and consumption. We report the parameter values for the stochastic processes for the baseline model in Table 1. The key parameters of the stochastic process are the serial correlation of the labor tax shock $\rho_l = .95$ and the ratio of the variance of the labor tax innovation to that of the technology shock innovation, $\sigma_l^2/\sigma_z^2 = 1.09$. We consider several other versions of our maximum likelihood procedure in our sensitivity analysis below. (In one of them we show that if we choose the measurement error to be one-thousandth of the variance of the corresponding variables in Y_t, we obtain similar parameter estimates. In others we consider different variables in the observer equation.)

The model's *impulse response of hours* to a technology shock is calculated recursively as follows. Start at a steady state, set the technology shock innovations $\varepsilon_{m0}^z = \Delta > 0$, $\varepsilon_{mt}^z = 0$ for $t \geq 1$, and set the labor tax innovations $\varepsilon_{mt}^d = 0$ for all t. From (9), (10), and (13) we see that the impact effect, namely the response in period 0, is $\psi_z\Delta$, the effect in period 1 is $\psi_k\gamma_z\Delta$, the effect in period $t \geq 2$ is $\gamma_k^{t-1}\psi_k\gamma_z\Delta$, and so on.

In Figure 1 we plot the model's impulse response. We see that on impact, a shock to technology leads to a persistent increase in hours. The left vertical axis follows the business cycle literature and measures the response to a 1% shock to total factor productivity, while the right vertical axis follows the SVAR literature and converts this response to a 1 standard deviation shock in ε_{mt}^z. In order to relate the two vertical axes, note that 1 unit on the left vertical axis corresponds to $(1 - \theta)\sigma_z$ units on the right axis. All our impulse response plots use two axes of this form. We see that on impact a 1% shock to total factor productivity leads to an increase in hours of .44%.

3. Testing the SVAR Procedures and Deconstructing Their Failure

We test the SVAR procedures by comparing the model's impulse responses to the population impulse responses from the SVAR procedures applied to data from the model. We focus mainly on population impulse responses rather than impulse responses computed from short samples for two reasons. First, at least since Hurwicz (1950), estimates of autoregressive processes have been known to be biased in small samples. Second, we wish to abstract from any issues regarding sampling uncertainty.

We nest both the DSVAR and the LSVAR procedures by considering the nested SVAR specification $X_t = (\Delta(y_t/l_t), l_t - \alpha l_{t-1})$, where α is between 0 and 1. Note that $\alpha = 1$ corresponds to the DSVAR specification and $\alpha = 0$ to the LSVAR specification. With α near 1, this formulation also nests a quasi-differenced specification, referred to as the QDSVAR specification, which we discuss below.

Figure 2 displays the model's impulse response and the (population) impulse responses for the DSVAR and the LSVAR procedures. We follow the literature in setting the lag length p equal to 4 in (1). This figure shows that both of the SVAR impulse responses are very different from the model's impulse response. For example, the impact coefficient from the DSVAR is both negative and large. The impact coefficient from the LSVAR is almost three times as large as that of the model and would be extremely difficult to reconcile with essentially any quantitative business cycle model. In this sense, both SVAR procedures fail our test in a quantitatively large way.

We turn now to deconstructing the source of this failure. In addition to the main identifying assumptions, the SVAR procedures make a key auxiliary assumption: in the models of interest, X_t has an autoregressive representation that is well approximated with a small number of lags, typically four. 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

15

representations, but the associated autoregressive representations are not well approximated with a small number of lags. Here we argue that the heart of the SVAR's failure is that the autoregressive representation is not well approximated with a small number of lags.

The DSVAR specification has one technical difficulty. For our model, specifying the SVAR in the difference in hours amounts to overdifferencing 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 technical difficulty is inessential in our findings, we consider a QDSVAR specification with α close to 1. As long as α is less than 1, showing that our model has an invertible moving average representation is easy. 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. (Note 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), Quah (1990), and Fernandez-Villaverde, Rubio, and Sargent (2005).)

We now show that the autoregressive representation of X_t is not well approximated with a small number of lags. To do so, consider the moving average representation of X_t implied by the model. Let this representation of X_t be denoted as

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

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 form

(15) $X_t = v_{mt} + C_{m1}v_{mt-1} + C_{m2}v_{mt-2} + \ldots,$

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

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

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. Proofs for the following and all subsequent propositions are in the technical appendix (Chari, Kehoe, and McGrattan (2005)).

Proposition 1. (Infinite Order AR) The autoregressive representation of X_t is infinite order and satisfies $B_{mi} = MB_{mi-1}$ for $i \geq 2$, where the decay matrix $M = C_{m2}(C_{m1})^{-1} - C_{m1}$, and has eigenvalues equal to the quasi-differencing parameter $\lambda_1 = \alpha$ and $\lambda_2 = (\gamma_k - \gamma_l\psi_k/\psi_l - \theta)/(1 - \theta)$, where $\gamma_k, \gamma_l, \psi_k$, and ψ_l are the coefficients of the decision rules in (10) and (13) and θ is the capital share.

Given our parameters for the QDSVAR specification (including the quasi-differencing parameter $\alpha = .99$), the eigenvalues are $\lambda_1 = .99$ and $\lambda_2 = .96$. For the LSVAR specification the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = .96$. Since the rate of decay is, at least asymptotically, determined by the largest eigenvalue, these eigenvalues suggest that an autoregression with a small number of lags is a poor approximation to the infinite order autoregression.

For illustrative purposes we consider an autoregression with one lag for which we can obtain an analytical expression for the impulse response coefficients. (Note that, quantitatively, the one-lag VAR and the four-lag VAR give similar impulse responses.) This autoregression is of the form

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

17

Let $\bar{C} = \sum_{i=0}^{\infty} C_i$ be the associated sum of the moving average coefficients. Recall that for an autoregression of the form (1) with an arbitrary number of lags, (5) implies that the impact coefficients depend only on the covariance matrix and the sum of the moving average coefficients. The following proposition relates the population estimates for the covariance matrix, Ω, and the sum of the moving average coefficients, $\bar{C}$, derived from a regression with one lag of the form (17), to the covariance matrix Ω_m of the innovations of X_t and the sum of the moving average coefficients in the model $\bar{C}_m$, associated with (15) and (16).

Proposition 2. (Error in Population Estimates from Specification Error) Given X_t from our model, the population estimate of the covariance matrix in (17) is

$$(18) \quad \Omega = \Omega_m + M \left(\Omega_m - \Omega_m V(X_t)^{-1} \Omega_m \right) M'$$

and the (inverse of the) sum of the moving average coefficients associated with (17) is

$$(19) \quad \bar{C}^{-1} = (\bar{C}_m)^{-1} + M(I - M)^{-1} C_{m1} + M(\Omega_m - V(X_t))V(X_t)^{-1},$$

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 impact coefficients for a technology shock can then be computed using (5). Since the Ω and $\bar{C}$ derived from the one-lag VAR generally differ from the model's Ω_m and $\bar{C}_m$, the impulse responses will typically differ as well. (Of course, since our model has an invertible moving average representation, 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 of the model's impulse response. Figure 3 displays the QDSVAR responses for lag lengths p ranging from 4 to 300. Notice that even with 20 lags, the QDSVAR procedure gives a poor approximation to the model's impulse response. From these figures 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.

Figure 4 shows the impulse responses from the LSVAR for lag lengths p ranging from 4 to 100. As with the QDSVAR, we see that the impulse response from the LSVAR is a good approximation to the model only for extremely long lag lengths. Accurately estimating VARs with such long lag lengths is infeasible in practice.

We argue that the need for a large number of lags when we run the VAR stems from the presence of capital in our model and the absence of capital in the VAR specification as well as from the presence of multiple shocks in our model. One way to understand the role of capital in the failure of the SVARs to uncover the model's impulse response is to drive the capital share to zero in the model. In this case, even when the SVARs have only one lag, the model's impulse response to a technology shock coincides with the population impulse response of the SVARs.

When only the technology shock is present, so that the variance of the labor tax τ_l is zero, $M\Omega_m M' = M\Omega_m V(X_t)^{-1}\Omega_m M'$. From (18) we then have that the covariance Ω in (17) is equal to that in the model Ω_m. Interestingly, because Ω_m is singular in this case, the two impulse responses coincide even though the moving average coefficients from the one-lag specification C_i do not equal the model's moving average coefficients C_{mi}.

We formalize this discussion in the following proposition.

19

Proposition 3. (Two Special Cases) When either the capital share θ is zero or the variance of the labor tax innovation σ_τ is zero, the SVAR procedures with one lag recover the model's impulse response.

Proposition 3 considers the limiting case when the model has no capital or no variability in labor taxes. We now consider the ability of the SVAR to uncover the model's impulse response away from the limit as we vary the capital share. In the next section we consider the SVAR's impulse response as we vary the stochastic process governing the shocks.

We focus on the specification error for the impact coefficient, defined as the percent error in the SVAR impact coefficient relative to that of the model. Figure 5A displays the specification error for the impact coefficient for the QDSVAR and the LSVAR specifications. We see that as the capital share decreases from its baseline value of .33, so does the specification error. Even with a capital share of only .2, the specification error is about 100% for both specifications.

In order to capture the shape of the impulse response, we also report on the half-life of the impulse response for the LSVAR specification. Figure 5B shows that the half-life for the LSVAR impulse response is typically less than that for the model, indicating that the LSVAR response decays more rapidly than the model's response. (Both here and later, for the most part, we do not report the half-life statistics for the QDSVAR impulse response; the impulse response coefficients are typically negative, so the half-life statistic is less interesting than it is for the LSVAR.)

We also experimented with increasing the depreciation rate as another way of reducing the importance of capital. We found that when the depreciation was so high that capital essentially depreciated completely within a year, the model and the SVAR impulse responses were close.

4. Sensitivity Analysis to the Stochastic Process

We next investigate the importance of the stochastic process governing shocks for the performance of the SVAR procedures. We report on how our findings change as we change the key parameters for our stochastic process for the shocks: the serial correlation parameter ρ_l and the ratio of the innovation variances σ_l^2/σ_z^2. (Note that since equilibrium of the model is computed by linearizing the first order conditions, only the ratio of the innovation variances matters for the impulse responses to a 1% shock.)

Figure 6A displays the specification error for the impact coefficient for the QDSVAR and the LSVAR specifications as we vary the ratio of the labor tax innovation variance to the technology shock innovation variance. We see that as the relative variability of the labor tax declines from its baseline value of 1.09, so does the specification error. In this sense, for first generation models in which technology shocks play a dominant role, the SVAR procedures perform relatively well. In terms of second generation models, however, note that even with a relative variability of half of its estimated value, the specification error is over 100% for both specifications. Figure 6B displays the half-life of the LSVAR impulse response as we vary the innovation variance ratio. This figure indicates that for a wide range of values of this ratio, the LSVAR impulse response decays more rapidly than the model's impulse response.

Figure 7A shows that for any given value of the innovation variance ratio, the specification error for the impact coefficient changes very little as we vary ρ_l from a relatively low value of .90 to a relatively high value of .99. The half-life of the impulse responses, however, is sensitive to ρ_l. Figure 7B displays the half-life of the LSVAR impulse response as we vary the innovation variance ratio for two values of the serial correlation parameter: the relatively low value of .90 and the relatively high value of .99. The figure shows that as the serial correlation rises, so does the half-life.

Figures 6 and 7 show that the SVAR impulse responses depend on the parameters of the stochastic process governing the labor tax shocks. Of course, in the model the impulse

21

response to a technology shock is independent of the parameters governing the labor tax shock. The dependence of the SVAR impulse responses arises because the SVAR is misspecified with respect to the number of lags and therefore confounds the effects of the technology shock and the labor tax shock.

We also investigated the sensitivity of our findings to three other ways of estimating the model. In the *low measurement error* version of the model, we set the measurement error of each variable at one-thousandth of the variance of the corresponding variables in Y_t. We estimated the key parameters to be $\rho_l = .95$ and $\sigma_l^2/\sigma_z^2 = 1.06$. Note that these parameters are close to those in the baseline model and, as shown by Figures 6 and 7, so are the results.

In the *restricted observer* version of the model, we let the observable variables be $Y_t = (\Delta \log y_t, \log l_t)'$, and since there are two variables in the observer equation and two shocks in the model, we set the measurement error to zero. Here the key parameters are given by $\rho_l = .998$ and $\sigma_l^2/\sigma_z^2 = .48$. For this version of the model, the specification errors for the impact coefficient of the QDSVAR and the LSVAR are 77% and 64%, respectively. The shape of the responses is quite different from that of the model in that, relative to the model, the QDSVAR decays very rapidly and the LSVAR decays very slowly. For example, the half-lives of the impulse responses are 6 and 158 quarters, respectively, compared to a half-life in the model of about 18 quarters.

In the *government consumption* model, we add a third orthogonal shock. This shock is normalized government spending $\hat{g}_t = g_t/Z_t$, which follows an autoregressive process of the form

$$(20) \quad \log \hat{g}_{t+1} = (1 - \rho_g) \log \bar{g} + \rho_g \log \hat{g}_t + \sigma_g \varepsilon_{mt+1}^g,$$

where ε_{mt}^g, together with our earlier shocks ε_{mt}^z and ε_{mt}^d, are jointly normal, independent of each other, and i.i.d. over time. For this version, we reestimated the parameter values for

the technology process and the labor tax process when we estimated the parameter values for the government spending process. We set the measurement error to be one-hundredth of the variance in the observed variables. The key parameter estimates are $\rho_l = .92$, $\sigma_l^2/\sigma_z^2 = 1.04$, $\rho_g = .97$, and $\sigma_g = .02$. For this version of the model, the specification errors for the impact coefficient of the bivariate QDSVAR and the bivariate LSVAR are –293% and 162%, respectively. The associated half-lives are 67.5 quarters and 9.8 quarters, compared to a half-life in the model of about 18 quarters.

We experimented with a variety of other ways to estimate the stochastic processes and found results similar to those reported here: the specification error for the impact coefficient was very high, or the half-life was very far off from that in the model, or both.

5. Adding Other Variables

So far we have focused on the SVARs with two variables: the log difference of 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 the effects of adding other variables to the SVARs. We show that when we add stochastic processes estimated from the data, our results do not change significantly. We go on to show that under extreme circumstances, in which the variances of the added shocks are much smaller than they are in our estimates, the impulse response from the SVAR procedure is approximately the same as the model's impulse response. We argue that the ability of the SVAR procedure to uncover the model's impulse response is related to the number of singularities in the decay matrix M and the covariance matrix of innovations Ω_m for the model.

Our discussion of Proposition 3 suggests that one of the problems with the SVAR specification is that it does not include a capital-like variable. In the model, the relevant state variable is $\hat{k}_t = k_t/Z_{t-1}$. Since Z_{t-1} is not observable, we cannot include $\hat{k}_t$ itself. We consider

23

several stationary capital-like variables: the capital-output ratio k_t/y_t, the investment-output ratio x_t/y_t, and the growth rate in the capital stock $\log k_{t+1} - \log k_t$. One conjecture is that including such variables 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, in general, incorrect.

As we show in the technical appendix (Chari, Kehoe, and McGrattan (2005)), when we add the capital-output ratio or the growth rate of the capital stock to the list of variables in the VAR, we find that the model's moving average representation of these variables is not invertible. In both specifications the autoregressive coefficients B_{mi} in (16) derived from the moving average coefficients of the model have decay matrix M, similar to that in Proposition 1. When we add the capital-output ratio one of the eigenvalues of M is $-\infty$, while when we add the growth rate of the capital stock one of the eigenvalues is 1. Since both specifications suffer from the type of invertibility problems discussed in Hansen and Sargent (1991), we do not investigate them here.

In practice most researchers prefer using the investment-output ratio as a capital-like variable rather than measures that use the capital stock directly because they feel that the capital stock is poorly measured. The combination of invertibility and measurement issues leads us to use the investment-output ratio to capture the influence of the capital-like variable. We find, however, that adding such a variable does not diminish the need for long lags in the VAR.

A. Same Number of Shocks and Variables

Consider a SVAR with three variables and three shocks. The third variable is the log of the investment-output ratio x_t/y_t, where $x_t = (1+\gamma)k_{t+1} - (1-\delta)k_t$. The third shock is normalized government spending $\hat{g}_t = g_t/Z_t$, which follows an autoregressive process of the form in (20). We use maximum likelihood to estimate the parameters of the stochastic

process. To make comparisons between the SVARs in this model and those in the baseline model easier, in this estimation we set the rest of the parameters, including those of the other stochastic processes, as in the baseline model. We refer to this version as the *baseline model with government consumption* and report the estimated parameters in Table 1. We also considered the government consumption model described above in which we estimated all of the parameters of the stochastic process.

We focus on the population impulse responses of the SVARs. Figure 8 compares the three-shock model's impulse response for labor against the population impulse response from a three-variable LSVAR with $X_t = (\Delta(y_t/l_t), l_t, x_t/y_t)$ and with four lags. Clearly, the LSVAR procedure does not uncover the model's impulse response. (We found that the three-variable QDSVAR procedure on the model with three shocks also does not uncover the model's impulse response.) For comparison purposes we also repeat the plot of the impulse response for the two-variable LSVAR from Figure 2. We see that adding x_t/y_t does not significantly improve the performance of the LSVAR. We found similar results for the government consumption model.

In the *baseline model with investment tax*, we let the investment tax be the third shock. We assumed that taxes on investment follow the autoregressive process

$$(21) \quad \tau_{xt+1} = (1 - \rho_x)\bar{\tau}_x + \rho_x \tau_{xt} + \sigma_x \varepsilon^x_{mt+1},$$

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. We estimate the parameters of (21) using maximum likelihood. Again, in this estimation we set the rest of the parameters as in the baseline model. We report the estimated parameters in Table 1. Figure 8 displays the resulting population impulse response for this three-shock model from a three-variable LSVAR with $X_t = (\Delta(y_t/l_t), l_t, x_t/y_t)$ and with four lags. Clearly, the LSVAR procedure does not uncover

25

the model's impulse response.

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

Proposition 4. Consider a three-shock model with either government spending or investment taxes as the third shock. In both cases, the autoregressive representation of X_t is infinite order and satisfies $B_{mi} = MB_{mi-1}$ for $i \geq 2$, where the decay 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. Interestingly, the largest eigenvalue of the decay matrix M is roughly the same in the two- and the three-variable SVARs. In this sense, adding a variable such as the investment-output ratio does not reduce the need for long lags.

We also experimented with four-variable SVARs and four shocks. Relative to the baseline model, we added shocks to government consumption and the tax on investment. In the SVAR specifications, we added the investment-output ratio and the consumption-output ratio as variables. The analog of Proposition 4 applies to this case with the four eigenvalues of the decay matrix given by $\alpha, (1 - \delta)/(1 + g_y), 0$, and 0. In terms of the impulse responses, we found that adding these variables did not significantly improve the ability of the SVARs to reproduce the model's responses to technology shocks.

We then examined the state space representation, similar to (12) and (14), of the three-shock model and asked if we could find a third variable for which the SVAR specification mimics the state space representation. In the LSVAR specification we found that if we added

k_{t+1}/y_t, namely, the ratio of the capital stock in period $t+1$ to output in period t, the SVAR representation mimics the state space representation. In this exceptional case, the LSVAR procedure uncovers the model's impulse response. This finding does not imply that adding k_{t+1}/y_t is a general prescription for success for the SVAR procedure. For example, when we add k_{t+1}/y_t to the QDSVAR specification, the SVAR representation does not mimic the state space representation and the SVAR procedure does not uncover the model's impulse response.

More generally, it may well be true that across models a careful examination of the state space representation for each model could lead to a different SVAR specification for each model which mimics the state space representation. If so, it seems both safer and more transparent to estimate the state space representation directly as we discuss below.

In sum, we have shown that, other than in one exceptional case, when stochastic processes on shocks are estimated using data, adding variables and an equal number of shocks does not improve the performance of the SVAR procedure.

B. More Variables than Shocks: The Singularity Rule

One question is whether any circumstances exist under which adding more variables to the SVAR will improve its performance. One possibility that Proposition 3 suggests is that the performance of the SVAR procedure will improve if we add more variables without adding more shocks. In this vein, we consider three-variable SVARs in which we add either the investment-output ratio or the consumption-output ratio with only two shocks—to technology and to the labor tax. We find that the population impulse response from the LSVAR procedure with one lag coincides with the model's impulse response, while that of the QDSVAR procedure does not. We also consider four-variable SVARs with both three shocks and two shocks. In both cases, the LSVAR procedure uncovers the model's impulse response. The DSVAR procedure uncovers the impulse response only in the two-shock case.

Our experiments suggest that the ability of the SVAR procedure to uncover the model's impulse response is related to the number of singularities in the system. The number of singularities is the sum of the number of zero eigenvalues of the decay matrix M in the AR representation, denoted $s(M)$, and the number of zero eigenvalues of the model's innovation covariance matrix Ω_m, denoted $s(\Omega_m)$. The following *singularity rule* holds in all of the examples we have considered with two or more shocks: the SVAR procedure, even with one lag, uncovers the model's impulse response if and only if the number of singularities is at least as large as the number of variables, so

$$s(M) + s(\Omega_m) \geq n,$$

where n is the number of variables in the VAR. When the model has only one shock, confounding the effects of various shocks is not possible and, under weak conditions, the SVAR procedure uncovers the model's impulse response.

We think of our singularity rule as a rule of thumb that holds in the class of business cycle models we have considered. An open question is whether similar rules hold in other classes of business cycle models. Clearly, our singularity rule does not apply to all models. For example, the SVAR procedure does not necessarily uncover the model's impulse response for models with invertibility problems of the type discussed in Hansen and Sargent (1991) even when such models satisfy our singularity rule.

In terms of applying our singularity rule, we note that, with one exception, in all the models with capital that we have examined, the decay matrix M has at least one nonzero eigenvalue, so $s(M) \leq n - 1$. Hence, in such models if the number of shocks is equal to the number of variables, so Ω_m has no singularities, the SVAR procedure with a small number of lags will not uncover the model's impulse response. (In the exceptional case when we add k_{t+1}/y_t to the VAR, it turns out that in the LSVAR specification, the decay matrix M has

all zero eigenvalues and, as our singularity rule implies, the LSVAR procedure uncovers the model's impulse response.)

Our singularity rule suggests that, absent exceptional cases, the SVAR procedure in practice is unlikely to uncover the model's impulse response. To see why, note that in the data, estimated covariance matrices Ω from VARs, as in (17), are never singular. It follows from the singularity rule that if the estimated covariance matrix Ω equals the model's innovation covariance matrix Ω_m, the SVAR procedure will uncover the model's impulse response only for exceptional cases in which $s(M) = n$.

One might interpret our singularity rule as suggesting that the SVAR procedure will approximately uncover the model's impulse response as long as a relatively small number of shocks (or "factors") accounts for the bulk of fluctuations in the data. This reasoning presumes that in this case, the estimated covariance matrix Ω from a VAR would be approximately singular.

We show that this presumption is faulty. In our estimated models with three or four shocks, we find that the added shocks contribute little to the fluctuations in variables such as output, labor, and investment. For example, fluctuations in government consumption (measured as the sum of government spending and net exports) account for about 3.4% of the fluctuations in the growth rate of output. Even though the shocks are small in this sense, the resulting SVAR does not perform well because in order for the SVAR procedure to uncover the impulse responses, the variances of the added shocks must be extraordinarily small.

For the three-variable LSVAR procedure with three shocks, we computed the LSVAR specification error for the impact coefficient, namely, the percentage difference between the population impulse response for the LSVAR and the model impulse response. At the estimated shock variances for government consumption and the investment tax, the specification errors are over 100% and close to 200%, respectively. Figure 9A displays the specification

29

error for the impact coefficient, defined as above, against the variance of the third shock for both the government consumption case and the investment tax case. Even for small shock variances, the specification error is large. For example, when the shock variances are one-half their estimated values, the specification error is still over 75% in both cases. Figure 9B displays the half-life of the LSVAR impulse responses as we change the variance of the third shock. The figure indicates that for a large range of values of these variances, the LSVAR impulse response decays more rapidly than the model's. In sum, in order for the SVAR procedure to perform well, the shock variances must be much smaller than we find them to be in the data.

We have shown that even when the contributions to the added shocks to fluctuations are as small as they are in the data, the estimated VAR covariance matrix Ω is sufficiently far away from being singular that the SVAR procedure does not perform well. Hence, while the conditions under which the SVAR procedure uncovers the model's impulse response are mathematically intriguing, we find them to be of little applied interest.

6. Small Sample Issues

Thus far, we have focused on population moments in order to abstract from small sample issues. We have shown that the specification error in the lag length leads to quantitatively large differences between the SVAR impulse responses in the population and the model's impulse responses. Here, in order to address small sample bias, sampling uncertainty, and the ability of standard tests to detect the need for long lags, we examine the SVAR procedures for samples of the same length as in the U.S. data.

We decompose the *overall difference* between the model's impulse response and the mean across short samples of the SVAR impulse responses into two parts. One part, the *specification error*, is the difference between the population SVAR impulse responses (with four lags) and the model's impulse responses. The other part, the *small sample bias,* is

30

the difference between the population SVAR impulse responses and the mean across short samples of the SVAR impulse responses.

A. The QDSVAR Procedure in Small Samples

Our implementation of the QDSVAR procedure in small samples is as follows. We treat the model as the data-generating process and draw 1,000 sequences of roughly the same length as our data length, namely, 180 quarters. We run the QDSVAR procedure on each sequence and report on features of the impulse responses of hours to technology shocks. The left panel of Figure 10A displays 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 10A 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. Figure 10B displays the mean impulse response across these 1,000 sequences as well as the mean of the bootstrapped confidence bands across the same sequences.

In comparing Figures 2 and 10B, we see that the small sample bias for the QDSVAR is quite small relative to the specification error. Hence, the overall difference between the model's impulse response and the short sample QDSVAR impulse response is large. We next ask, suppose for each of the 1,000 sequences a researcher tested the hypothesis that the impact coefficient of the QDSVAR 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 whether standard lag length tests uncover the need for more lags. For each simulation we compute the Akaike information statistic for each lag. We found the Akaike criterion selects a lag length of four or fewer for over 98% of the simulations. We

31

also found that a likelihood ratio test did not reject four lags in favor of five in 93% of the simulations. We experimented with other lag length tests and found similar results. Overall, we found that with samples of roughly the same length as U.S. data, standard lag tests do not uncover the need for long lags.

The reason the Akaike-like lag length tests do not detect the need for more lags is as follows. These tests balance the gain in the fit of the model from adding more parameters against a fixed penalty for doing so. As we add more parameters, the gain in the fit of the model is smaller than the penalty.

We also implemented the QDSVAR procedure as follows. For each sequence we let a lag length criterion determine the appropriate lag length of the VAR and then compute impulse responses. Our results were unchanged.

Finally, we experimented with a *short-run* sign convention in which a technology shock is called positive if it raises the level of labor productivity on impact. We found that our results were the same.

B. The LSVAR Procedure in Small Samples

We implement the LSVAR procedure in the same fashion as we did the QDSVAR procedure.

Figures 11A and 11B, the analog of Figures 10A and 10B, indicate that the range of impulse response coefficients is very wide. For example, 95% of the impact coefficients lie between $-.45$ and 1.94. Figure 11A suggests that the LSVAR procedure is not useful for distinguishing between, for example, sticky price models and 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

32

that if monetary policy is not at all accommodative, hours fall in response to a technology shock. The range of responses for hours to a technology shock in sticky price models is well within our 95% range in Figure 11A.

In comparing Figure 11B to Figure 2, we see that the small sample bias for the LSVAR is sizeable and of the opposite sign as the specification error. Nevertheless, the overall difference between the model's impulse response and the mean of the short sample LSVAR impulse response is substantial. For example, for the impact coefficient, the mean of the short sample impulse is more than twice as large as that of the model.

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 zero at the 5% significance level. Such a researcher would essentially never reject this hypothesis either. These findings suggest that with the kind of data generated by our model, the LSVAR is incapable of differentiating between models with starkly different impulse response functions.

We conducted lag length tests parallel to those we conducted for the QDSVAR procedure. We found the Akaike criterion selects a lag length of four or fewer for over 99% of the simulations. We also found that a likelihood ratio test did not reject four lags in favor of five in 94% of the simulations. We experimented with other lag length tests and found similar results. As in the QDSVAR case, we found that with samples of the same length as U.S. data, standard lag tests do not uncover the need for long lags.

As in our analysis of the QDSVAR procedure, we also let a lag length criterion determine the appropriate lag length of the VAR and then computed impulse responses. Our results were unchanged.

Finally, we experimented with the short-run sign convention and found that the 95%

range analogous to that in Figure 11B widens significantly and includes many more negative values.

C. Putting Our Findings in Context

In order to put our findings in context, we apply the SVAR procedures to three popular U.S. data sets used in the SVAR literature: Francis and Ramey (2004), Christiano, Eichenbaum, and Vigfusson (2003), and Gali and Rabanal (2004). All three data sets use conceptually similar measures of productivity and hours. We find that when the quasi-differencing parameter is close to one, the procedure gives similar results: in all three data sets, a technology shock leads to a fall in hours on impact.

Here we focus on the impulse responses and the associated confidence bands obtained by running the LSVAR specification with four lags on these data sets. We find that the LSVAR procedure yields sharply differing results for the three data sets. On impact a technology shock leads to a fall in hours in one, leads to a rise in hours in another, and leaves hours roughly unchanged in the third. These large differences in findings across similar data sets are likely to be connected to our finding regarding the wide range of LSVAR impulse responses across simulations from our model.

For the first series we use the data that Francis and Ramey (2004) constructed to estimate an LSVAR for the period 1948:1–2002:4. Their measure of productivity is the Bureau of Labor Statistics (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. Figure 12A illustrates that an innovation resulting in a 1% 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.

For the second series, we follow Christiano, Eichenbaum, and Vigfusson (2003), who use the DRI Basic Economics 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). Figure 12B shows that a positive technology shock leads to a persistent rise in hours. On impact a 1% 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 three onward.[5]

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 the Haver USECON database with mnemonics for output, hours, and population of LXNFO, LXNFH, and LNN, respectively. Figure 12C indicates 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.

These sharply contrasting results lead researchers in the SVAR literature to draw sharply contrasting inferences. Francis and Ramey (2004) argue that this evidence shows that real business cycle models are dead. Christiano, Eichenbaum, and Vigfusson (2003) maintain that the models are alive and well. Gali and Rabanal (2004) assert that these results, by themselves, are inconclusive. Gali and Rabanal prefer the alternative DSVAR specification, which, they argue, shows that real business cycle models are also dead.

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 dif-

---

[5]Christiano, Eichenbaum, and Vigfusson (2003) use an instrumental variables procedure that Shapiro and Watson (1988) proposed, rather than our OLS procedure, and they compute Bayesian confidence intervals rather than our bootstrapped confidence intervals. Comparing our Figure 11B with Figure 2 in their paper, we see that our results are essentially the same.

ferences in measuring productivity and hours raises some questions about the reliability of the LSVAR procedure for drawing inferences about underlying models.

7. State Space Approaches as Alternatives

We have pointed out difficulties with using SVAR procedures to guide the development of economic theory. Here we discuss an alternative state space approach that is safer and more transparent and uses more economic theory.

As we have shown, standard business cycles such as ours have the state transition equation $S_{t+1} = FS_t + G\varepsilon_{mt+1}$ and the observation equation $Y_t = HS_t + u_t$. The state transition equation and the observation equation form a standard linear state space system that is easy to estimate by standard statistical methods. The parameters of F, G, and H are all functions of the original structural parameters of the model. (In our baseline model, the structural parameters include those of preferences and technology, such as $\beta, \theta, \delta, \phi$, along with the parameters of the shock processes.) The standard procedure is to estimate the structural parameters using the cross-equation restrictions implied by the theory.

An alternative approach imposes many fewer restrictions than the theory. Under this approach, we estimate the (reduced form) parameters F, G, and H directly rather than the structural parameters. This approach requires that we impose only the minimal identifying assumptions to estimate F, G, and H. As long as these minimal identifying assumptions are consistent with the model generating the data, standard estimation procedures, such as maximum likelihood, will, in population, recover the state space parameters. Hence, the population impulse responses from the state space system will coincide with those from the model.

Consider applying this alternative approach to data from our baseline model. Specifically, consider using demeaned data from our baseline model to construct population estimates of the parameters of the state space system, F, G, and H. We let the state of the

36

system S_t equal $(\log \hat{k}_t, \log z_t, \tau_{lt}, \log \hat{k}_{t-1}, \tau_{lt-1})$ after subtracting their means. Showing that this state is minimal is easy. We choose the observed variables $Y_t = (\Delta \log(y_t/l_t), \log l_t)$ to be the same as those in the LSVAR specification. We normalize the units of all the shocks so that their variances are unity. We also normalize the coefficient on the response of the capital stock to a technology shock to be unity. This normalization is a choice of the units in which we measure $\hat{k}_t$. In our state space system, in addition to the normalizations we impose three other types of restrictions: the variables that enter the state of the system are stationary, the state is minimal, and the stochastic processes are the same as in our baseline model. We then have the following proposition.

Proposition 5. (Identification of the State Space Parameters) The parameters of the state space system are identified.

The key step in the proof of this proposition is to show that the similarity transformation conditions in Burmeister, Wall, and Hamilton (1986) that guarantee identification are satisfied in our case. Clearly, then, the population impulse responses obtained from maximum likelihood will coincide with the model's impulse responses. An open question is how well this procedure will work in small samples.

This state space procedure can be applied more broadly. In other applications, if we wanted to claim that the impulse responses from such a procedure coincided with those from a broad class of models, including models with sticky prices, models with financial frictions, and so on, we would have to find a state space representation and a set of identifying assumptions that nested the class of models of interest. Note that estimating the state space system is very similar to a number of alternative statistical strategies including estimating the appropriately specified VARMA specification.

An approach that is closely related to this state space approach but imposes more restrictions is the business cycle accounting approach of Chari, Kehoe, and McGrattan (2004).

8. Related Literature

In the literature, critiques of the SVAR approach are not new. They can be broadly 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, 1991) point out that invertibility problems may plague the type of Box–Jenkins methods that underlie the SVAR literature. (See also Fernandez-Villaverde, Rubio, and Sargent (2005).) They show 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), analyze 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 different from the Hansen–Sargent invertibility critique.

Cooley and Dwyer (1998) lucidly critique the SVAR procedure using economic models as tests in a manner broadly similar to ours. One important difference between our work and theirs, however, is that they mainly focus on models that violate the key assumptions of the SVAR approach either by not having a unit root in the technology shock or by having correlated shocks. We focus on models that satisfy the key assumptions of the SVAR approach and show through a series of propositions that even then the SVAR approach may fail to uncover the model's impulse responses. 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, whereas they focus on a variety of other issues. See also McGrattan (2004) for work

similar in spirit to that of Cooley and Dwyer.

Erceg, Guerrieri, and Gust (2004) also test the SVAR procedure using economic models. In contrast to our focus on mistaken inferences using theoretical propositions about population moments, their main focus is on small sample bias in SVARs, and they conclude that the small sample bias problem in models is modest. Most important, they conclude, "Overall, Gali's methodology appears to offer a fruitful approach to uncovering the results of technology shocks"(p. 4). 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 "unless strong restrictions are applied, conventional inferences regarding impulse responses will be badly biased in all sample sizes" (p. 345). 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 shows that in infinite-dimensional spaces, unless severe restrictions are imposed on the parameters, standard methods cannot be used to make asymptotically valid confidence statements.

9. Conclusion

Simple data analysis techniques that reliably point us toward quantitatively promising models can be highly useful in applied economic analysis. The SVAR literature argues that the SVAR procedure is a robust technique for guiding the development of theory. Specifically, this literature claims that as long as the model satisfies the key identifying assumptions, the procedure will uncover the model's impulse response. We have provided counterexamples to this claim by showing that for a large set of parameters the procedure does not uncover the

model's impulse response. Furthermore, we show that this large set of parameters includes those estimated from data.

We have shown analytically that the SVAR procedure performs better the less important are the demand shocks. In this sense, the SVAR procedure performs relatively well in first generation models in which technology shocks account for the bulk of fluctuations in output. This procedure performs relatively poorly for second generation models that emphasize the role of other shocks in accounting for fluctuations in a broader set of aggregates. Since most current business cycle research is focused on developing second generation business cycle theories, our findings do not augur well for the ability of the SVAR procedures to guide the development of such theories.

Our study reinforces the point that Hansen and Sargent (1991) made 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 impossible to determine the answers to basic issues such as the ones discussed here: For what questions will a short lag length be reasonable? For what type of stochastic processes on the shocks will the procedure do better and for what type will it do worse? What variables should we include in the VAR?

We have shown that simply adding more variables to the VAR will not necessarily lead the SVAR procedure to uncover the model's impulse response. Typically, in the data, estimated covariance matrices from VARs are not singular, so models must have at least as many shocks as variables. In such situations, our singularity rule implies that the SVAR procedure will not uncover the model's impulse response. We have also shown that even if the variances of some of the shocks are much smaller than they are in the data, the SVAR procedure leads to mistaken conclusions.

A much safer and more transparent approach is to use economic theory to guide the specification. The state space approach, which uses more economic theory than that used in the SVAR procedures, is likely to be fruitful. This approach begins by recognizing that

business cycle models have state space representations and 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 parameters of the resulting state space model.

Elsewhere, in Chari, Kehoe, and McGrattan (2004), we have argued that the business cycle accounting approach, which is related to the state space approach but uses more theory, is useful in guiding the development of business cycle theories. This approach has the same goal as the SVAR approach—namely, to quickly shed light on which of a class of models is promising—but it suffers from fewer of the shortcomings.

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. In *Handbook of Computational Economics*, vol. 1, edited by Hans M. Amman, David A. Kendrick, and John Rust. Amsterdam: North-Holland/Elsevier.

Bencivenga, Valerie R. 1992. An Econometric Study of Hours and Output Variation with Preference Shocks. *International Economic Review* 33 (2): 449–71.

Benhabib, Jess, Rogerson, Richard, and Wright, Randall. 1991. Homework in Macroeconomics: Household Production and Aggregate Fluctuations. *Journal of Political Economy* 99 (6): 1166–87.

Bernanke, Ben S., Gertler, Mark, and Gilchrist, Simon. 1999. The Financial Accelerator in a Quantitative Business Cycle Framework. In *Handbook of Macroeconomics,* vol. 1C, edited by John B. Taylor and Michael Woodford. Amsterdam: North-Holland/Elsevier.

Blanchard, Olivier Jean, and Quah, Danny. 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review* 79 (4): 655–73.

_____. 1993. The Dynamic Effects of Aggregate Demand and Supply Disturbances: Reply. *American Economic Review* 83 (3): 653–58.

Braun, R. Anton. 1994. Tax Disturbances and Real Economic Activity in the Postwar United States. *Journal of Monetary Economics* 33 (3): 441–62.

Burmeister, Edwin, Wall, Kent D., and Hamilton, James D. 1986. Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering. *Journal of Business and Economic Statistics* 4 (2): 147–60.

Chari, V. V., Kehoe, Patrick J., and McGrattan, Ellen. 2004. Business Cycle Accounting. Staff Report 328, Federal Reserve Bank of Minneapolis.

_____. 2005. Technical Appendix to "A Critique of Structural VARs Using Real Business Cycle Theory" Working Paper, Federal Reserve Bank of Minneapolis.

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

Christiano, Lawrence J., Eichenbaum, Martin, and Vigfusson, Robert. 2003. What Happens After a Technology Shock? Working Paper 9819, National Bureau of Economic Research.

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

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

Cooley, Thomas F., and Hansen, Gary D. 1989. The Inflation Tax in a Real Business Cycle Model. *American Economic Review* 79 (4): 733–48.

Erceg, Christopher, Guerrieri, Luca, and Gust, Christopher. 2004. Can Long-Run Restrictions Identify Technology Shocks? International Finance Discussion Paper 792, Board of Governors of the Federal Reserve System.

Faust, Jon, and Leeper, Eric M. 1997. When Do Long-Run Identifying Restrictions Give Reliable Results? *Journal of Business and Economic Statistics* 15 (3): 345–53.

Fernandez-Villaverde, Jesus, Rubio, Juan, and Sargent, Thomas. 2005. A, B, C, (and D)'s for Understanding VARs. Mimeo, New York University.

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.

_____. 2004. A New Measure of Hours Per-Capita with Implications for the Technology-Hours Debate. Mimeo, University of California, San Diego.

Gali, Jordi. 1999. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? *American Economic Review* 89 (1): 249–71.

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: 723–43.

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

Greenwood, Jeremy, and Hercowitz, Zvi. 1991. The Allocation of Capital and Time over the Business Cycle. *Journal of Political Economy* 99 (6): 1188–214.

Hall, Robert E. 1997. Macroeconomic Fluctuations and the Allocation of Time. *Journal of Labor Economics* 15 (1): 223–50.

Hansen, Gary D. 1985. Indivisible Labor and the Business Cycle. *Journal of Monetary Economics* 16 (3): 309–27.

Hansen, Lars Peter, and Sargent, Thomas. 1980. Formulating and Estimating Dynamic Linear Rational Expectations Models. *Journal of Economic Dynamics and Control* 2 (1): 7–46.

_____. 1991. Two Difficulties in Interpreting Vector Autoregressions. In *Rational Expectations Econometrics*. Underground Classics in Economics. Boulder and Oxford: Westview Press.

Hurwicz, Leonid. 1950. Least Squares Bias in Time Series. In *Statistical Inference in Dynamic Economic Models,* edited by Tjalling C. Koopmans. New York: Wiley.

Kydland, Finn E., and Prescott, Edward C. 1982. Time to Build and Aggregate Fluctuations. *Econometrica* 50 (6): 1345–70.

Lippi, Marco, and Reichlin, Lucrezia. 1993. The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment. *American Economic Review* 83 (3): 644–52.

Marcet, Albert. 2004. On the Invertibility of "Non-invertible" Processes: Implications for VARs in Differences. Mimeo, Universitat Pompeu Fabra.

McGrattan, Ellen. 1994. The Macroeconomic Effects of Distortionary Taxation. *Jour-

nal of Monetary Economics 33 (3): 573–601.

—————. 2004. Comment on Gali and Rabanal's "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data?" *NBER Macroeconomics Annual 2004,* vol. 19*,* 289–308, edited by Mark Gertler and Kenneth Rogoff. Cambridge, Mass.: MIT Press.

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

Rotemberg, Julio J., and Woodford, Michael. 1992. Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity. *Journal of Political Economy* 100 (6): 1153–1207.

Shapiro, Matthew D., and Watson, Mark. Sources of Business Cycle Fluctuations. *NBER Macroeconomics Annual 1988,* vol. 3, 111–48, edited by Stanley Fischer. Cambridge, Mass.: MIT Press.

Sims, Christopher. 1971. Distributed Lag Estimation When the Parameter Space Is Explicitly Infinite-Dimensional. *Annals of Mathematical Statistics* 42: 1622–36.

—————. 1972. The Role of Approximate Prior Restrictions in Distributed Lag Estimation. *Journal of American Statistical Association* 67: 169–75.

—————. 1989. Models and Their Uses. *American Journal of Agricultural Economics* 71: 489–94.

Stockman, Alan C., and Tesar, Linda L. 1995. Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements. *American Economic Review* 85 (1): 168–85.

Uhlig, Harald. 2004. What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. Mimeo, Humboldt University.

TABLE 1

PARAMETERS OF THE STOCHASTIC PROCESSES

FOR BASELINE MODEL ^{a}

$$\mu_z = 0.00516 \ (0.000333), \ \ \sigma_z = 0.0131 \ (0.000598)$$

$$\bar{\tau}_l = 0.243 \ (0.00199), \ \rho_l = 0.952 \ (0.000955), \ \ \sigma_l = 0.0136 \ (0.000598)$$

FOR BASELINE MODEL WITH GOVERNMENT CONSUMPTION ^{b}

$$\bar{g} = 0.17, \ \rho_g = 0.969 \ (0.00244), \ \ \sigma_g = 0.0206 \ (0.00106)$$

FOR BASELINE MODEL WITH INVESTMENT TAX ^{c}

$$\bar{\tau}_x = 0.3, \ \rho_x = 0.98, \ \sigma_x = 0.0123 \ (0.00146)$$

Note: Parameters were estimated using maximum likelihood with quarterly data on output, hours, investment, and government consumption for the period 1959:1–2004:3. Numbers in parentheses are standard errors. Sources of basic data are the Bureau of Economic Analysis (National Income and Product Accounts) and the Bureaus of Labor Statistics and the Census (Current Population Survey).

[a] Parameters for technology and the labor tax used in our bivariate examples were estimated jointly, holding government consumption and the investment tax constant. For our trivariate examples, we hold estimates of technology and labor tax processes fixed at the values reported in this table and conditionally estimate parameters of either government consumption or the tax on investment.

[b] The parameter $\bar{g}$ was set so that the steady state share of government consumption was 20%.

[c] Parameter $\bar{\tau}_x$ is not identified and was fixed at its reported value. To ensure stationarity of the tax on investment, we had to constrain the parameter ρ_x.

FIGURE 1

Model Impulse Response of Hours to a Technology Shock
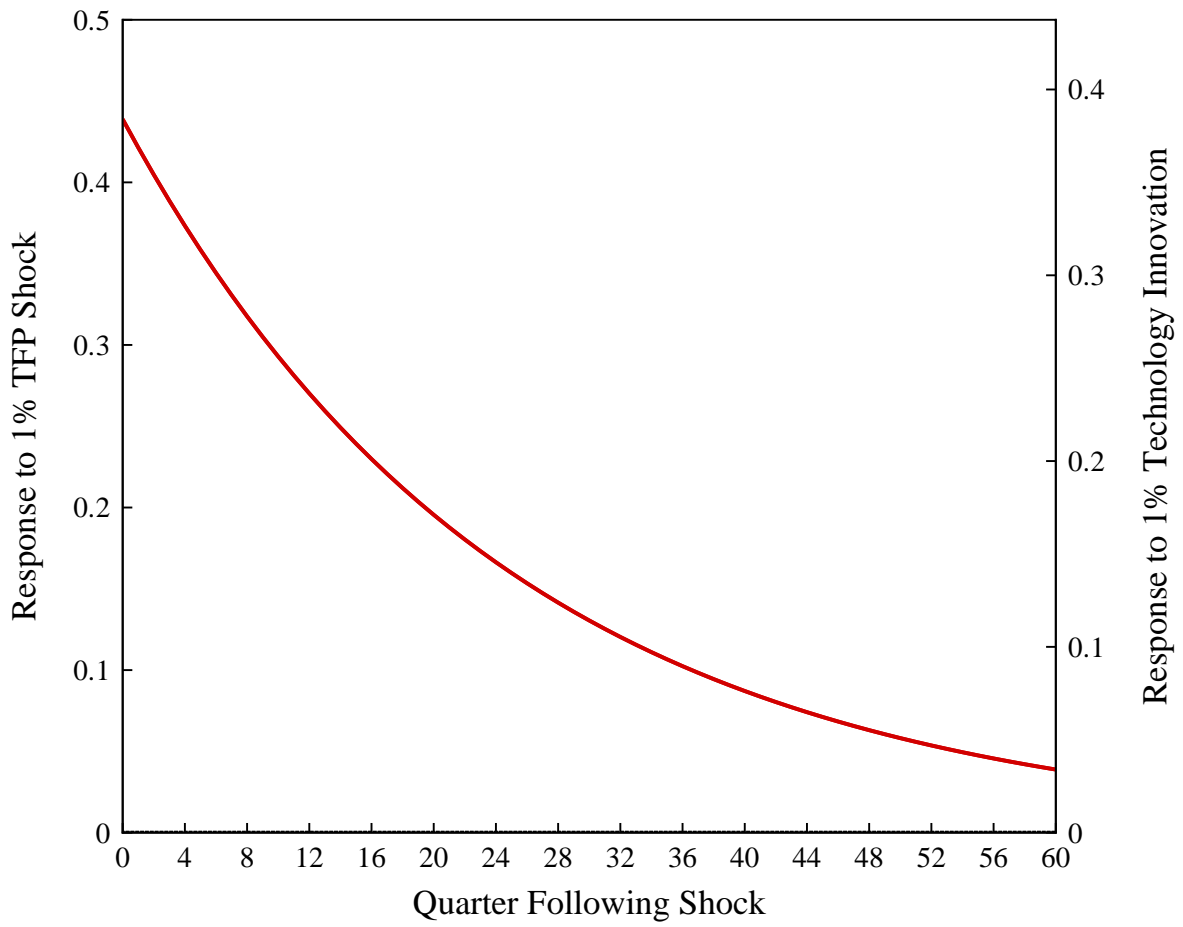
FIGURE 2

Impulse Responses of Hours for the Model and Population Responses
for the DSVAR and LSVAR Procedures with Four AR Lags
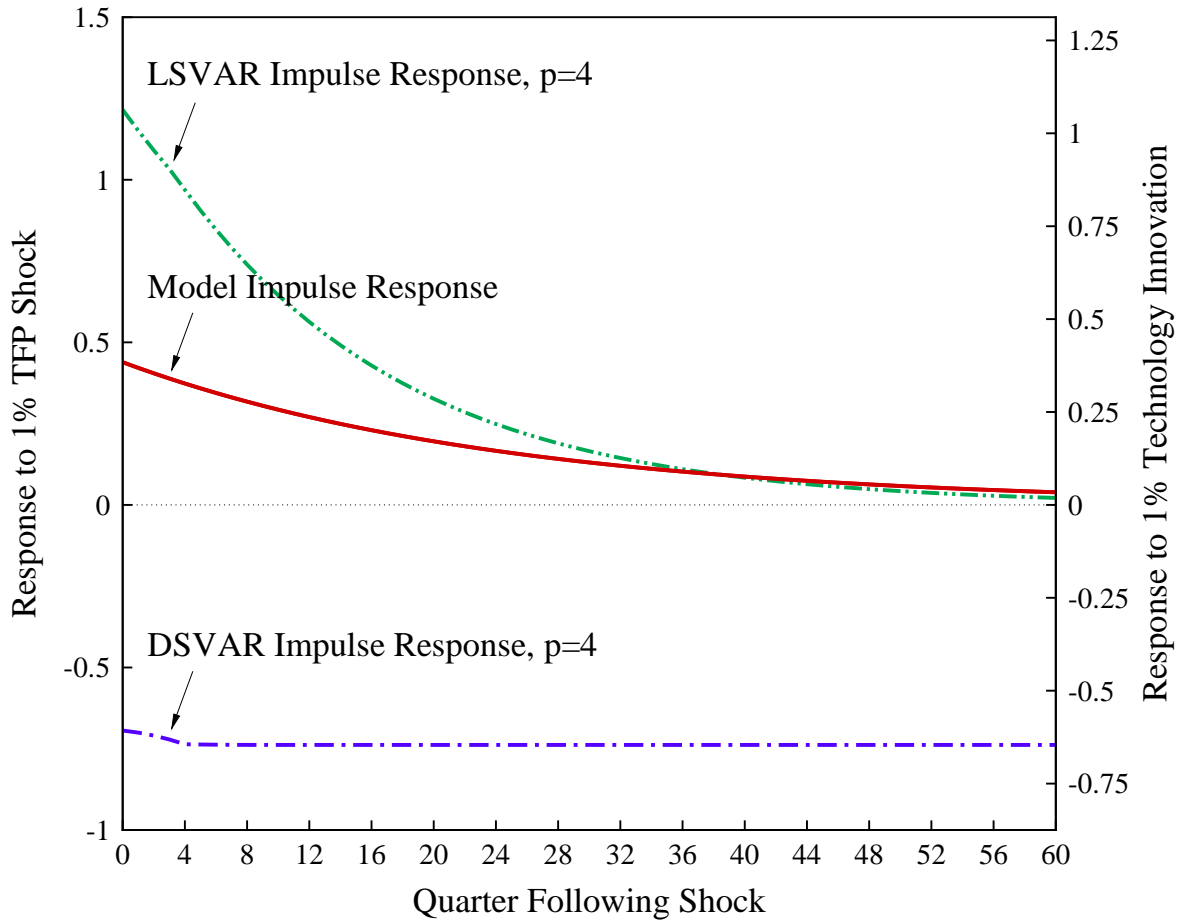
FIGURE 3

Impulse Responses of Hours for the Model and Population
Responses for the QDSVAR Procedure with Various AR Lags
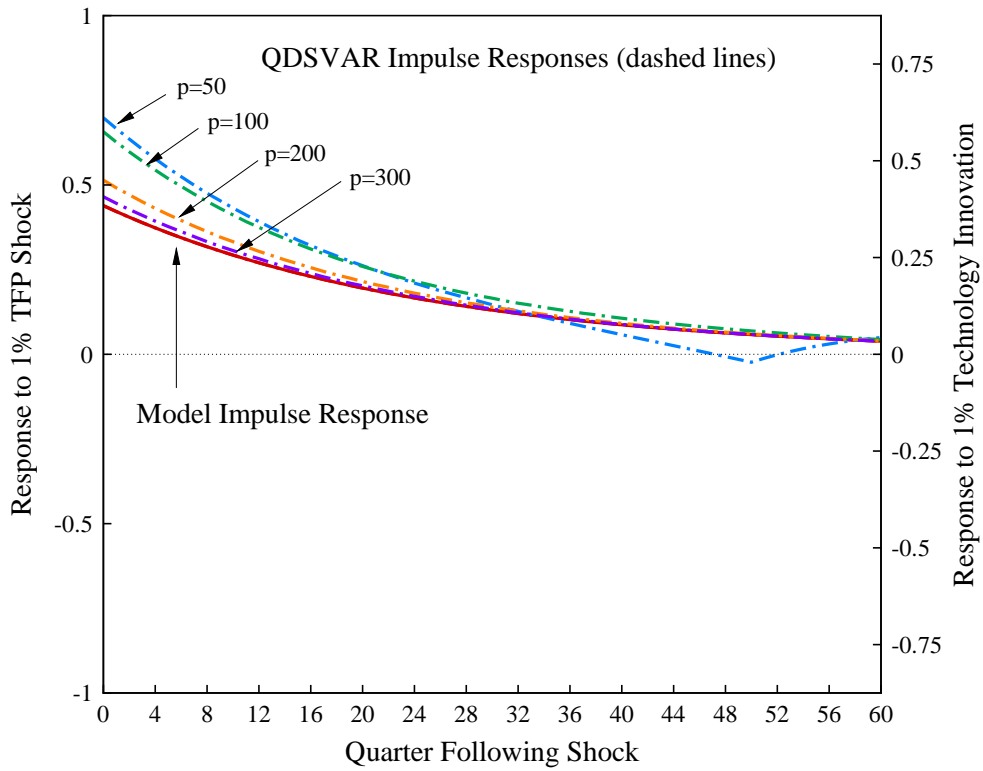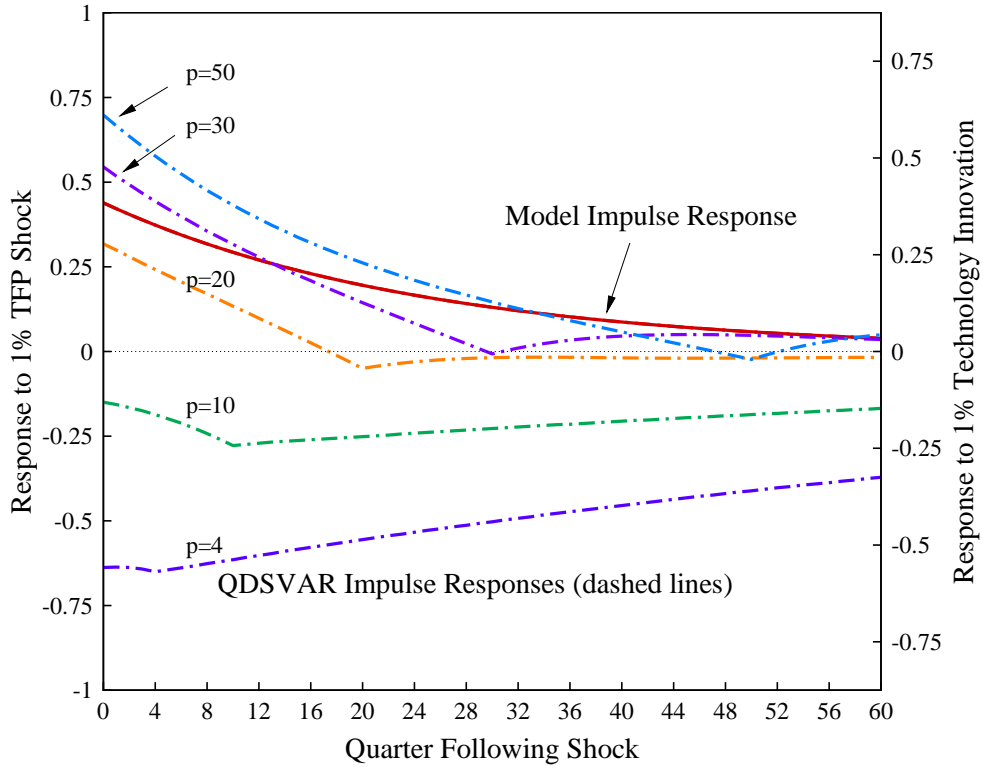
FIGURE 4

Impulse Responses of Hours for the Model and Population
Responses for LSVAR Procedure with Various AR Lags
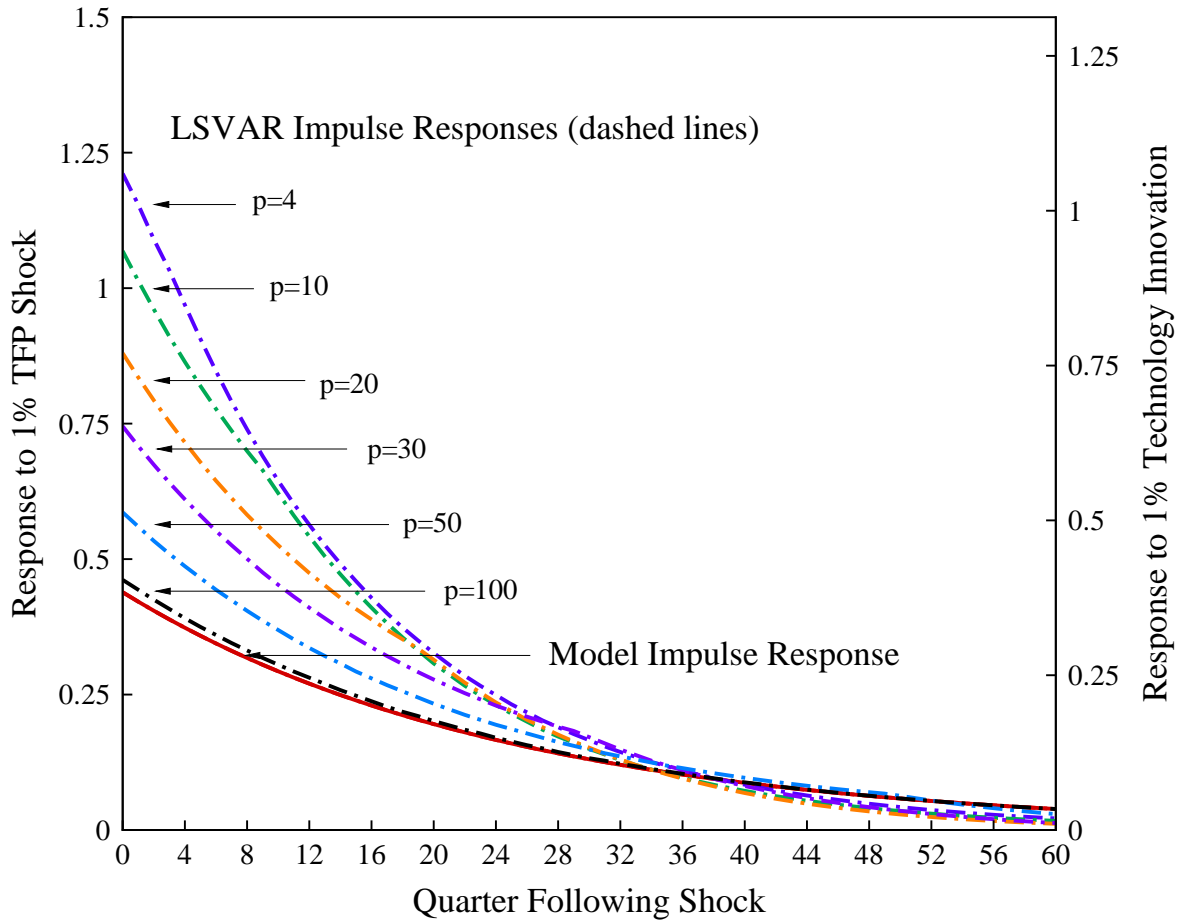
FIGURE 5A

Specification Error in the Impact Coefficient for the Four-Lag
SVAR Procedures, Varying the Capital Share
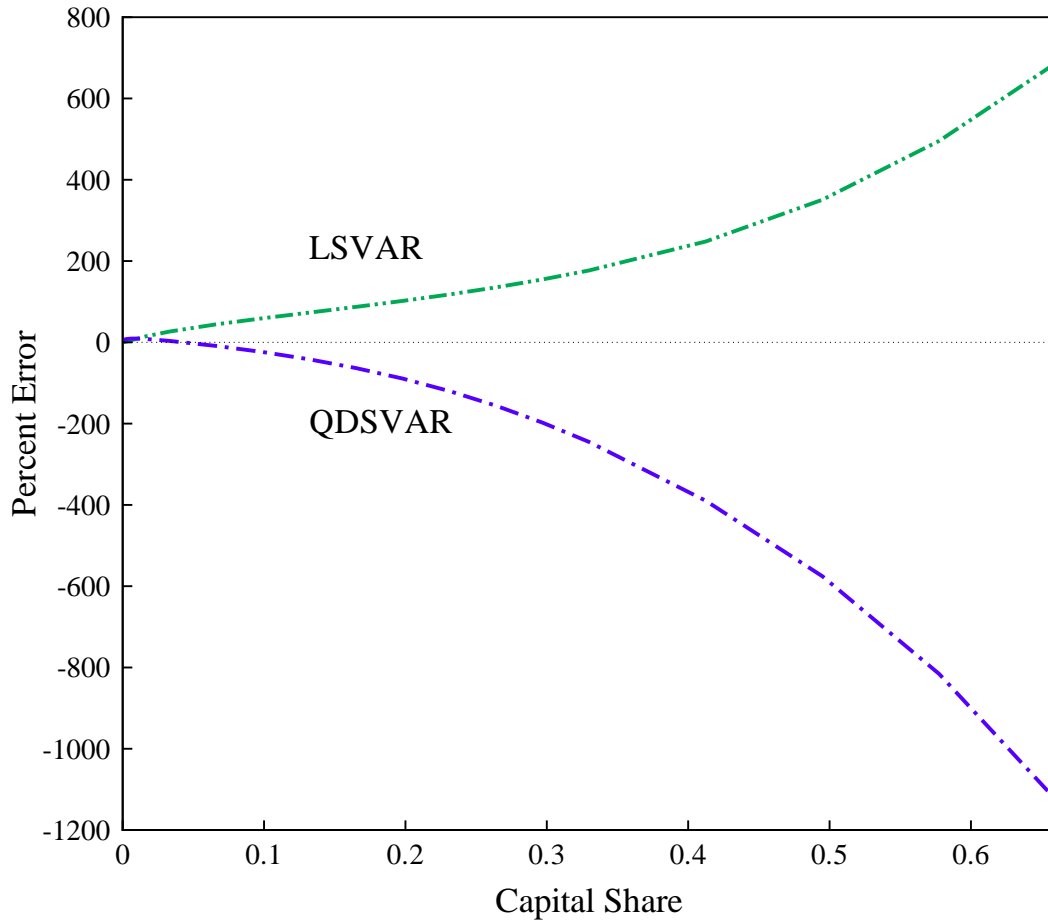
NOTE: The specification error is defined to be the percent error in the SVAR response
of hours to technology on impact relative to the model's theoretical response.

Half-Lives of the Impulse Responses for the Four-Lag
LSVAR Procedure, Varying the Capital Share
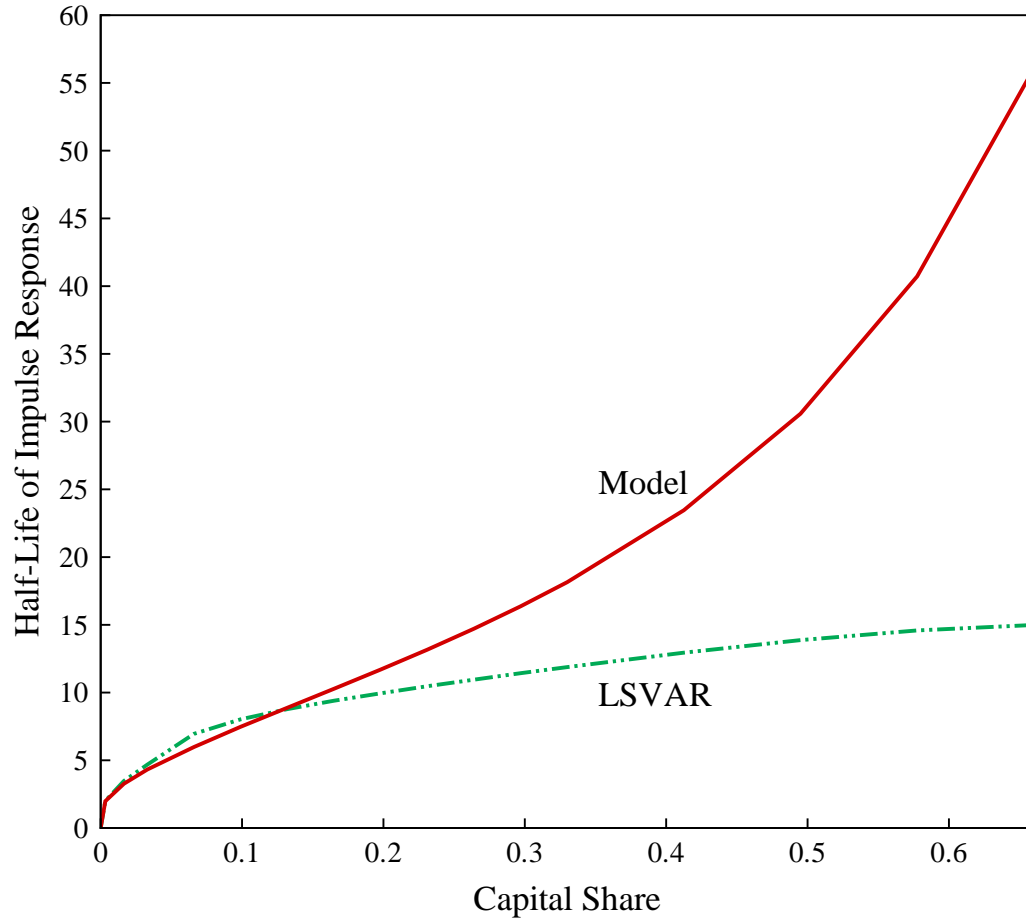
Specification Error in the Impact Coefficient for the Four-Lag
SVAR Procedures, Varying the Ratio of Innovation Variances

NOTE: The specification error is defined to be the percent error in the SVAR response
of hours to technology on impact relative to the model's theoretical response.

Half-Lives of Impulse Responses for the Four-Lag
LSVAR Procedure, Varying the Ratio of Innovation Variances
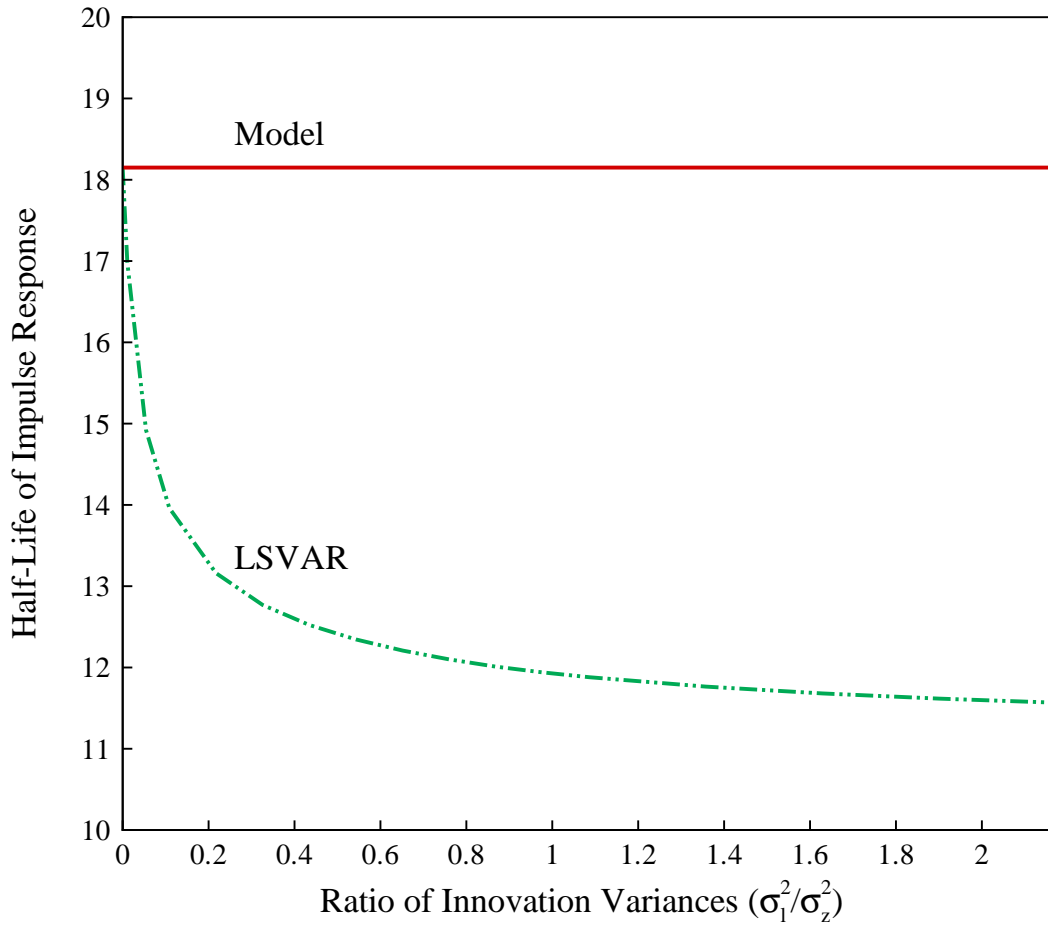
Specification Error in the Impact Coefficient for the Four-Lag
SVAR Procedures, Varying the Ratio of Innovation Variances and
the Serial Correlation of the Labor Tax Rate

NOTE: The specification error is defined to be the percent error in the SVAR response
of hours to technology on impact relative to the model's theoretical response.

Half-Lives of Impulse Responses for the Four-Lag LSVAR
Procedure, Varying the Ratio of Innovation Variances and
the Serial Correlation of the Labor Tax Rate
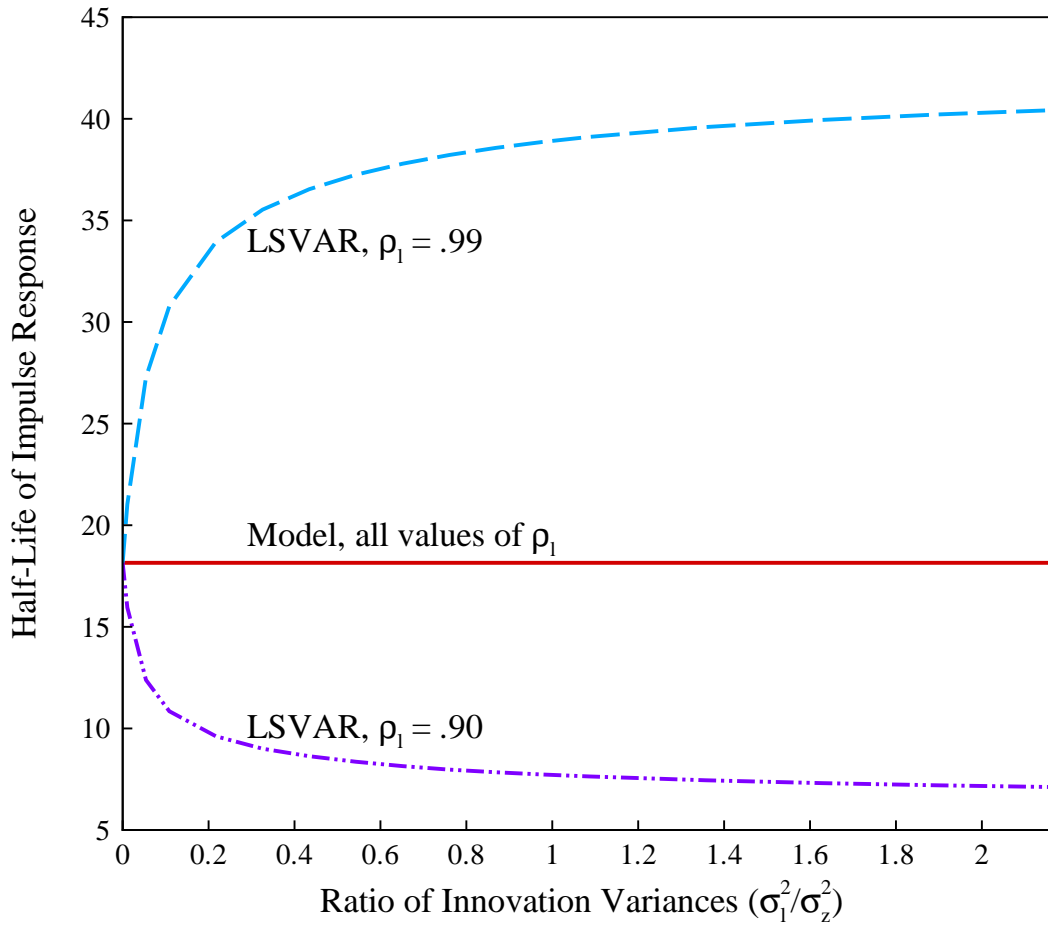
FIGURE 8

Impulse Responses of Hours for the Model and Population Responses
for Two- and Three-Variable, Four-Lag LSVAR Procedures

LSVAR Impulse Responses, p=4

3-variable VAR with τ_x shock

2-variable VAR

3-variable VAR with g shock

Model Impulse Response

Response to 1% TFP Shock

Response to 1% Technology Innovation

Quarter Following Shock
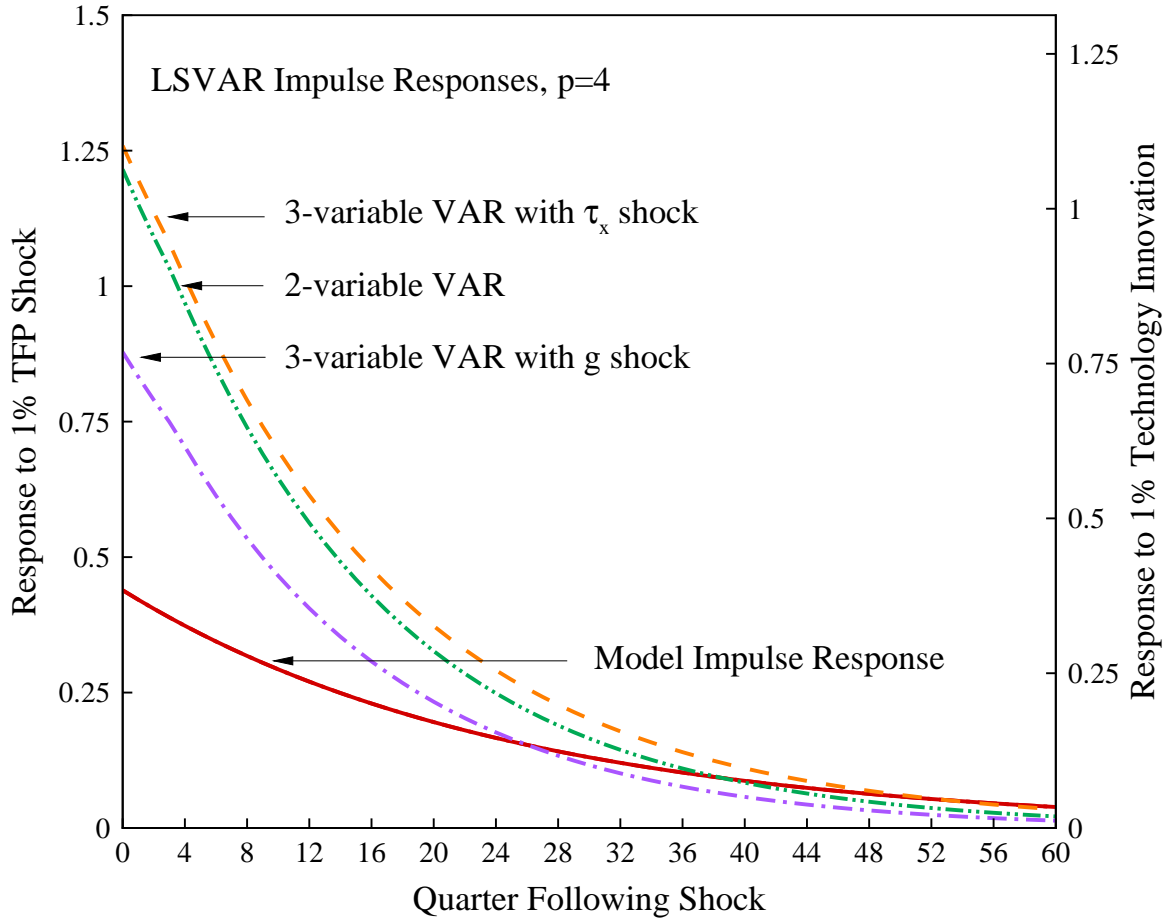
Specification Error in the Impact Coefficient for the Three-Variable,
Four-Lag LSVAR Procedure, Varying the Variance of the Third Innovation
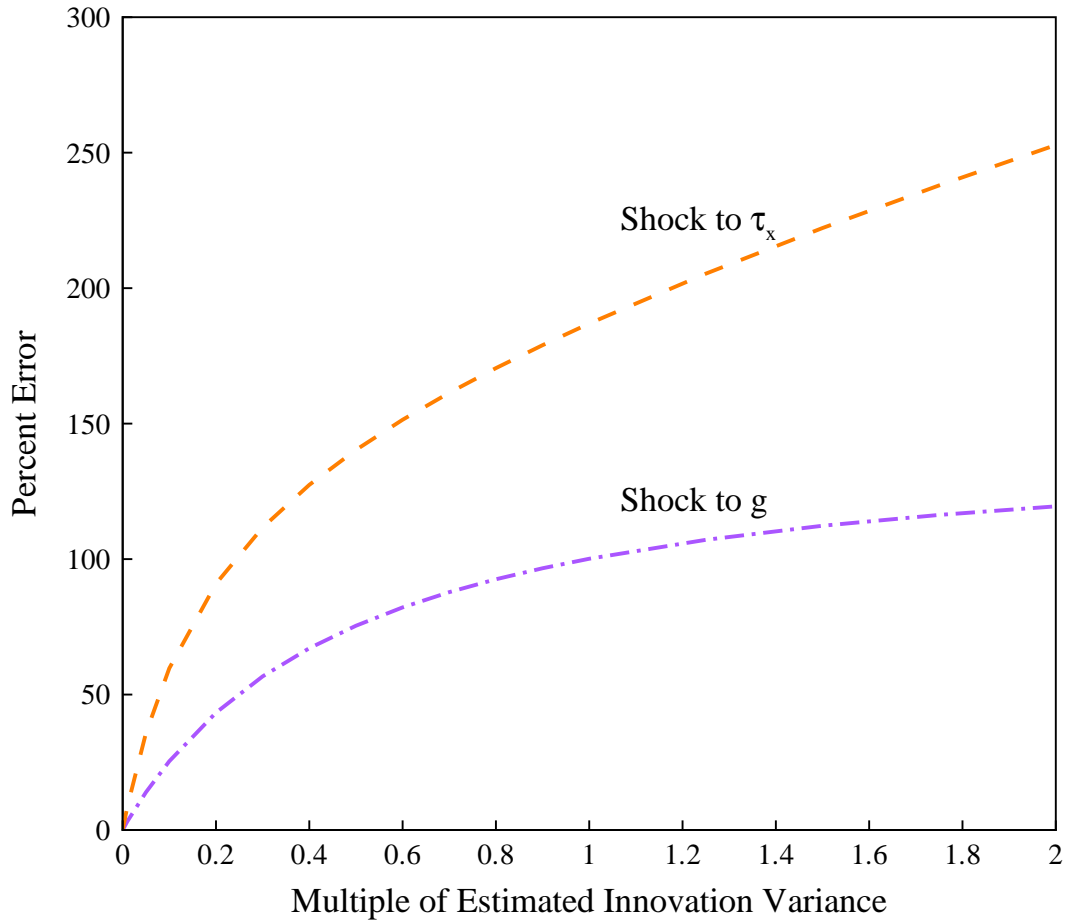
Half-Lives of Impulse Responses for the Three-Variable, Four-Lag
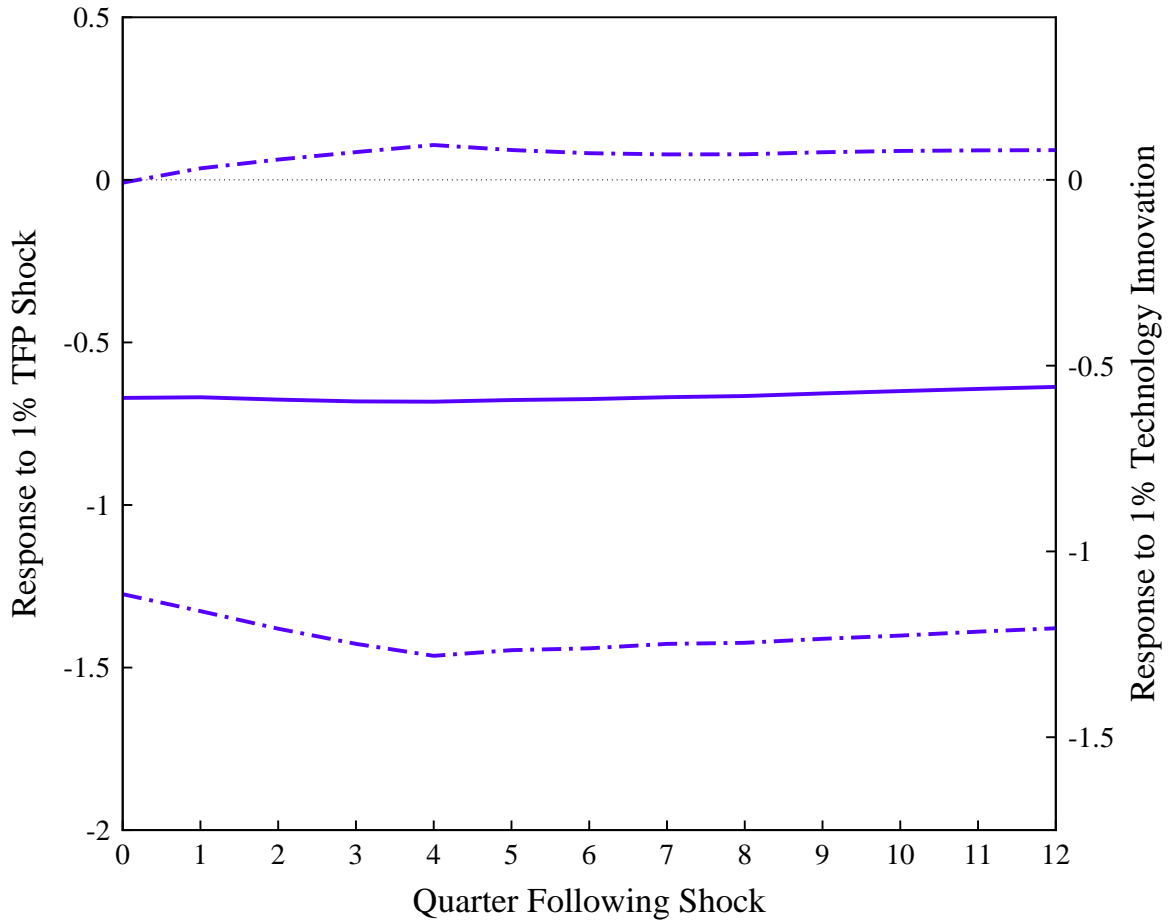LSVAR Procedure, Varying the Variance of the Third Innovation
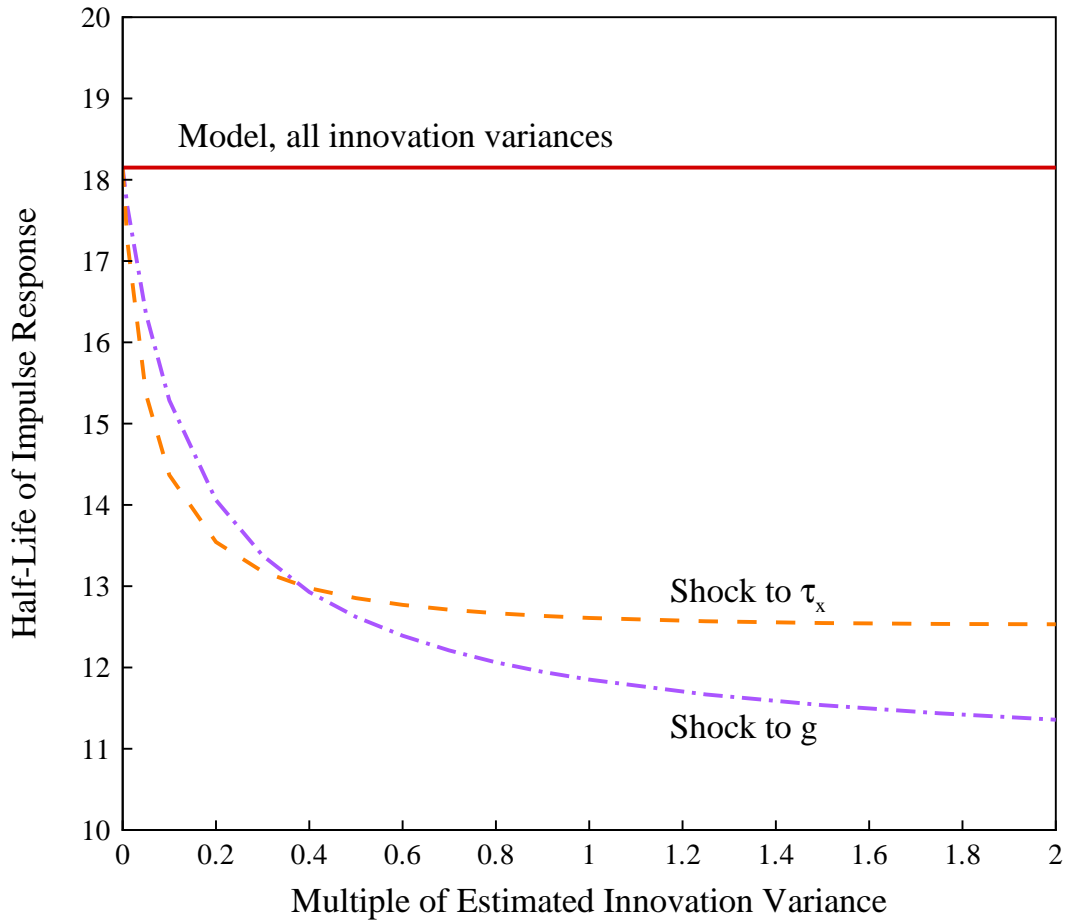
Histogram of Initial Impulse Response of Hours and 95% Bounds on
Responses Across 1,000 Applications of Four-Lag QDSVAR Procedure
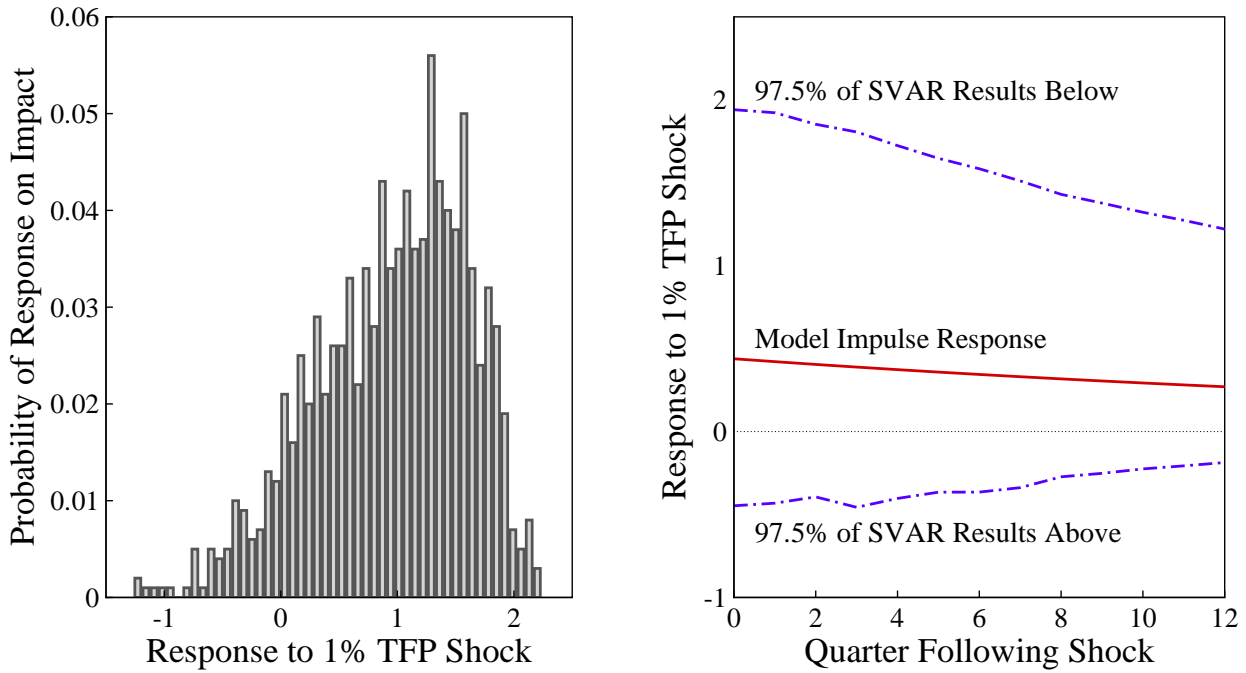to Model Simulations of Length 180
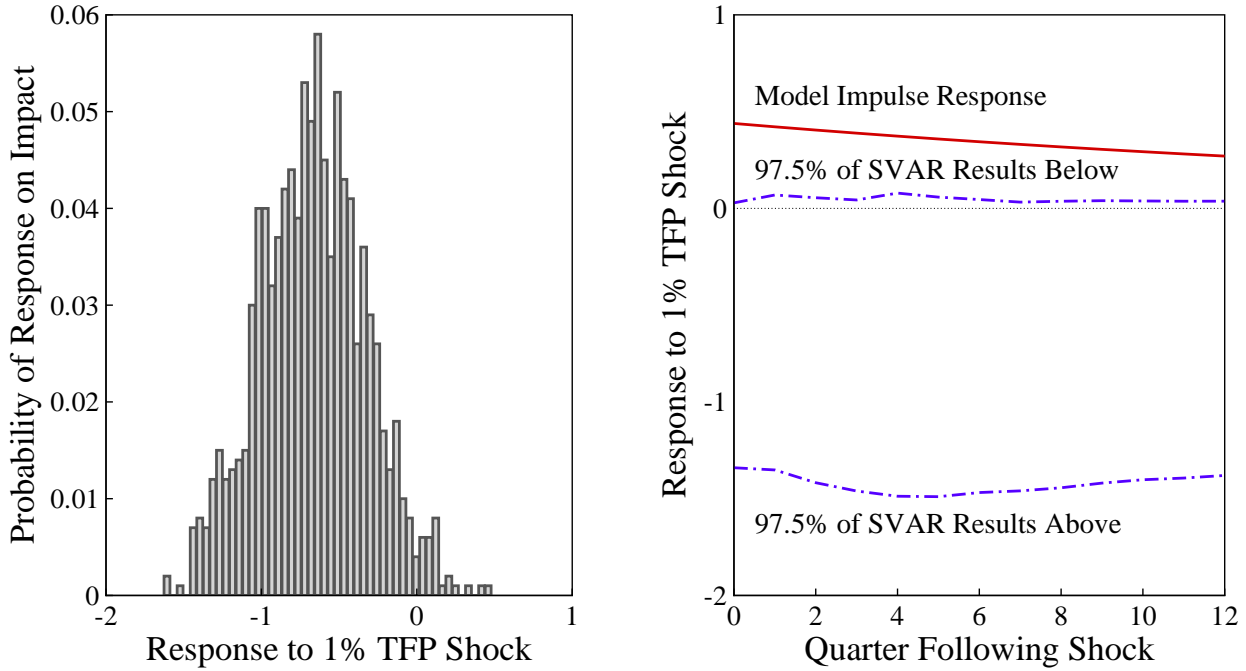
Mean Impulse Response of Hours (solid line) and Mean of 95% Bootstrapped
Confidence Bands (dashed lines) Averaged Across 1,000 Applications
of the Four-Lag QDSVAR Procedure to Model Simulations of Length 180

Histogram of Initial Impulse Response of Hours and 95% Bounds on
Responses Across 1,000 Applications of Four-Lag LSVAR Procedure
to Model Simulations of Length 180

Mean Impulse Response of Hours (solid line) and Mean of 95% Bootstrapped
Confidence Bands (dashed lines) Averaged Across 1,000 Applications
of the Four-Lag LSVAR Procedure to Model Simulations of Length 180
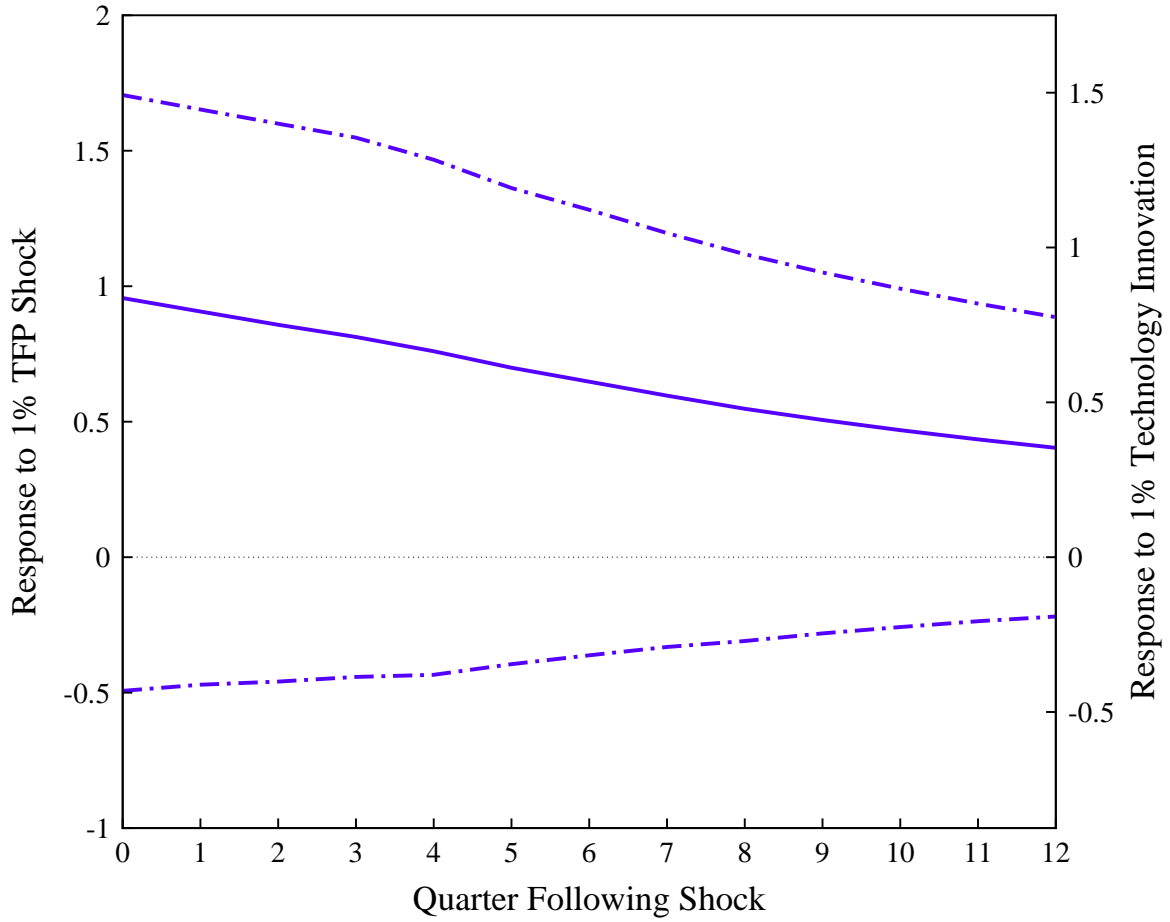
Impulse Response of Hours to a Technology Shock (solid line)
and Confidence Bands (dashed lines) Using the Four-Lag LSVAR
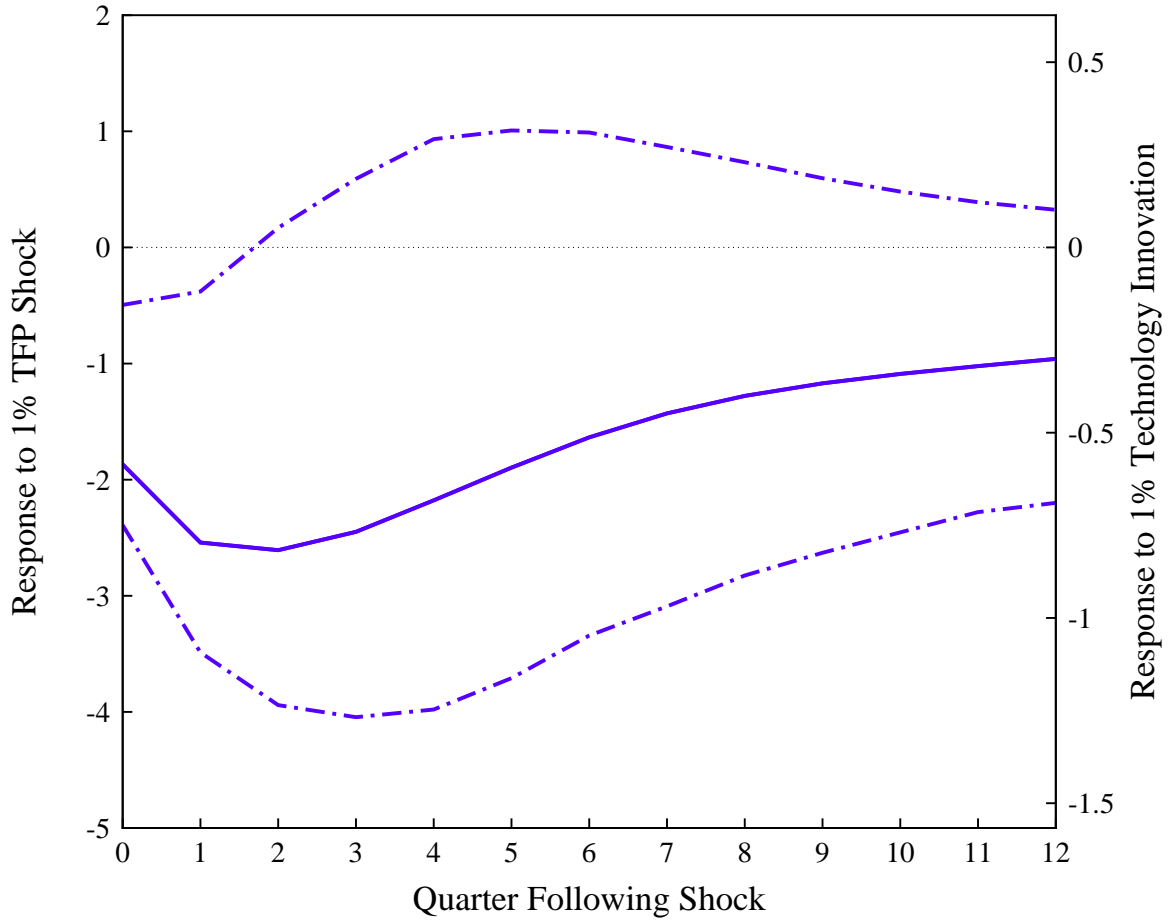Procedure with U.S. Data Set of Francis and Ramey (2004)

FIGURE 12B

Impulse Response of Hours to a Technology Shock (solid line)
and Confidence Bands (dashed lines) Using the Four-Lag LSVAR Procedure
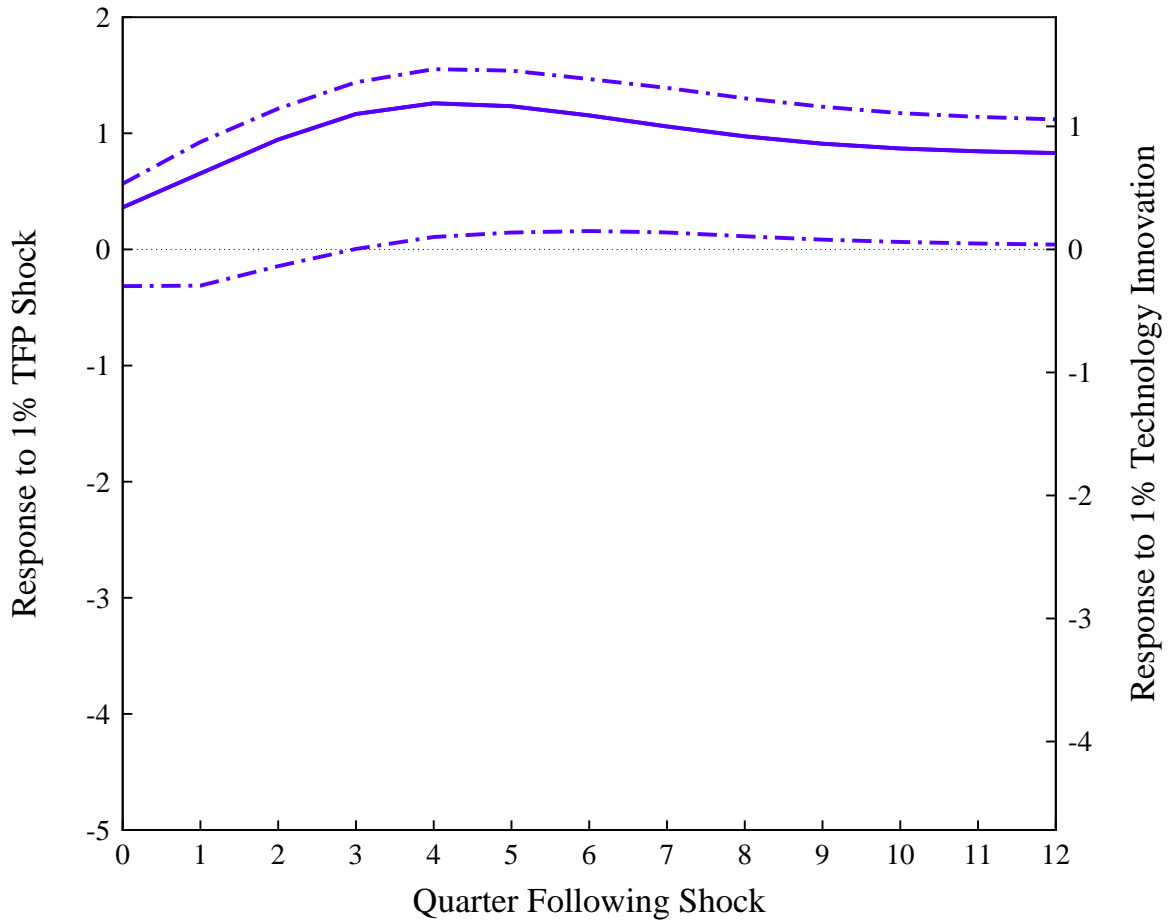with U.S. Data Set of Christiano, Eichenbaum, and Vigfusson (2003)

Impulse Response of Hours to a Technology Shock (solid line)
and Confidence Bands (dashed lines) Using the Four-Lag LSVAR
Procedure with U.S. Data Set of Gali and Rabanal (2004)