Federal Reserve Bank of Minneapolis
Research Department Staff Report 364

Revised May 2007

Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?*

V. V. Chari

University of Minnesota
and Federal Reserve Bank of Minneapolis

Patrick J. Kehoe

Federal Reserve Bank of Minneapolis
and University of Minnesota

Ellen R. McGrattan

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

The central finding of the recent structural vector autoregression (SVAR) literature with a differenced specification of hours is that technology shocks lead to a fall in hours. Researchers have used this finding to argue that real business cycle models are unpromising. We subject this SVAR specification to a natural economic test and show that when applied to data from a multiple-shock business cycle model, the procedure incorrectly concludes that the model could not have generated the data as long as demand shocks play a nontrivial role. We also test another popular specification, which uses the level of hours, and show that with nontrivial demand shocks, it cannot distinguish between real business cycle models and sticky price models. The crux of the problem for both SVAR specifications is that available data require a VAR with a small number of lags and such a VAR is a poor approximation to the model's VAR.

*The authors thank the National Science Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

The growing interest in structural vector autoregressions (SVARs) with long-run restrictions stems largely from the recent finding of researchers using this procedure that a technology shock leads to a fall in hours. Since a technology shock leads to a rise in hours in most real business cycle models, the researchers argue that their SVAR analyses doom existing real business cycle models and point to other types of models, such as sticky price models, as promising. (See Galí 1999, Francis and Ramey 2005a, and Galí and Rabanal 2005.) For example, Francis and Ramey write that "the original technology-driven real business cycle hypothesis does appear to be dead" (2005a, p. 1380) and that the recent SVAR results are "potential paradigm-shifters" (2005a, p. 1380). Similarly, Galí and Rabanal state that "the bulk of the evidence" they report "raises serious doubts about the importance of changes in aggregate technology as a significant (or, even more, a dominant) force behind business cycles" (2005, p. 274). We argue that these researchers' conclusions—and the usefulness of their procedure—are suspect when the procedure is closely examined.

In general, using SVARs to evaluate alternative economic models is an attempt to develop business cycle theory using a simple time series technique and minimal economic theory. In the *common approach* to this sort of analysis, researchers run VARs on the actual data, impose some identifying assumptions on the VARs in order to back out empirical impulse responses to various shocks, and then compare those empirical SVAR impulse responses to theoretical responses that have been generated by the economic model being evaluated. Models that generate theoretical responses that come close to the SVAR responses are thought to be promising, whereas others are not.

Here we focus on the SVAR literature that uses a version of this common approach with long-run restrictions in order to identify the effects of technology shocks on economic aggregates. The main claim of this literature is that its particular SVAR procedure can confidently distinguish between promising and unpromising classes of models without the researchers having to take a stand on the details of nontechnology shocks, other than minimal assumptions like orthogonality.

We evaluate this claim by subjecting the SVAR procedure to a natural economic test. We treat a multiple-shock business cycle model as the data-generating mechanism, apply the SVAR procedure to the model's data, and see if the procedure can do what is claimed for it.

We find that, in principle, the SVAR claim of not needing to specify the details of nontechnology shocks is correct if the researcher has extremely long time series to work with.

Regardless of the magnitude and persistence of other shocks, a researcher who applies the SVAR procedure to extremely long time series drawn from our model will conclude that the data are generated from our model and will be able to confidently distinguish whether the data are generated by our model or by a very different model.

With series of the length available in practice, however, the SVAR claim is incorrect. Our test shows that the impulse responses to technology shocks identified by the SVAR procedure vary significantly as the magnitude and persistence properties of other shocks vary, even though, obviously, the theoretical impulse responses do not. In particular, depending on the specification of the VAR, when other shocks play a nontrivial role in output fluctuations, a researcher who applies the SVAR procedure to data from our model either will conclude that the data are not generated from our model or will not be able to confidently distinguish whether the data are generated by our model or by a very different model. If, however, other shocks play only a trivial role in output fluctuations, then the SVAR impulse responses are close to the theoretical ones, and researchers can use the impulse responses to confidently distinguish between our model and very different models.

We obtain intuition for our findings from two propositions—an *infinite-order representation result* and a *first-order representation result*. The infinite-order representation result shows that when a VAR has the same number of variables as shocks, the variables in the VAR have an infinite-order autoregressive representation in which the autoregressive coefficients decay at a constant rate. Since we use a two-variable VAR and our model has two shocks, this result implies that the VAR has an infinite-order representation. With our parameter values, the coefficients in this representation decay very slowly. Even so, if very long time series are available, the empirical impulse responses are precisely estimated and close to the theoretical impulse responses.

With series of the length available in practice, however, the estimated impulse responses are not close to the theoretical impulse responses when the nontechnology shock is not trivial. A deconstruction of the SVAR's poor performance reveals that its problem is that the small number of lags in the estimated VAR dictated by available data lengths makes the estimated VAR a poor approximation to the infinite-order VAR of the observables from the model. That is, the VAR suffers from *lag-truncation bias*.

Our other proposition shows that, when the VAR has sufficiently many variables relative to the number of shocks, the VAR has a first-order representation.[1] This proposition implies that when only technology shocks are present, our two-variable VAR has such a representation. When nontechnology shocks play a sufficiently small role in generating output fluctuations, continuity implies that a VAR with two observables and a small number of lags well-approximates the true autoregressive representation. Hence, our first-order representation result suggests why when nontechnology shocks are small, the empirical and theoretical impulse responses are close.

Our test uses a stripped-down business cycle model which satisfies the key assumptions of the SVAR procedure. Researchers using this procedure make several assumptions in order to identify two types of underlying shocks, often labeled *demand shocks* and *technology shocks*. The two key identifying assumptions are that demand and technology shocks are orthogonal and that demand shocks have no permanent effect on the level of labor productivity, whereas technology shocks do—a common *long-run restriction*.

Our business cycle model also has two shocks, a technology shock and a demand shock, the latter of which resembles either a tax on labor income or a tax on investment, depending on the context. The business cycle model's technology shock is a unit root process, its demand shock is a first-order autoregressive process, and the two shocks are mutually independent. We show that the model satisfies the two key identifying assumptions of the SVAR procedure.

In implementing our test of that procedure, we need to take account of two quite different popular specifications. Both of these include two variables in their VAR: the growth rate of labor productivity and a form of hours worked. The differenced specification, or *DSVAR*, uses the first difference of hours, whereas the level specification, or *LSVAR*, uses the level of hours. In both specifications, because of the limited length of the available time series, the VAR is estimated with a small number of lags, typically four.

We sidestep one minor technical issue for one SVAR specification, the existence of an autoregressive representation of the model. The DSVAR specification does not have such a representation because hours worked are overdifferenced and the moving-average representation has a root of one, which is at the edge of the noninvertibility region of roots.[2] Instead

[1] Our first-order representation result suggests that simply adding enough variables to the VAR will ensure that the VAR procedure works well. Although this theoretical suggestion seems promising, we argue that it should be treated with caution if actual data are thought to have a large number of shocks relative to the number of observables that might typically be used in a VAR.

[2] One critique of the DSVAR procedure is that in all economic models, the time series hours worked per

of the DSVAR, therefore, we test here an alternative specification in which hours are quasi-differenced, called the *QDSVAR* specification. The variables in this specification do have an infinite-order autoregressive representation. And when the quasi-differencing parameter is close to one, the impulse responses of the QDSVAR and the DSVAR are indistinguishable.

We also ask which specification a researcher would prefer, the QDSVAR or the LSVAR, on a priori grounds. The time series of hours worked in our model is highly serially correlated, and we find that standard unit root tests do not reject the hypothesis that the hours series has a unit root. At least since Hurwicz (1950), we have known that autoregressions on highly serially correlated variables are biased in small samples and that quasi-differencing such variables may diminish that bias. Since both the QDSVAR and the LSVAR specifications have desirable asymptotic properties, on a priori grounds the QDSVAR seems preferable.

We test both of these SVAR specifications with the typical small number of lags. First we generate data from the business cycle model, drawing a large number of sequences of roughly the same length as postwar U.S. data. Then we run the two SVAR specifications with four lags on each sequence of model-generated data and compute the means of the impulse responses and the confidence bands.[3] Finally, we compare the SVAR impulse responses to those of the theoretical model, to see how well this procedure can reproduce the model's responses.

We find that contrary to the claim of the SVAR literature, the accuracy of the SVAR impulse responses depends critically on what type of shock has the most effect on output. When demand shocks account for a trivial fraction of output fluctuations, the means of the SVAR impulse responses are close to the model's theoretical impulse responses. When demand shocks account for a substantial fraction of output fluctuations, the SVAR means are very different from the model's theoretical impulse responses. Moreover, except when demand shocks account for a trivial fraction of output fluctuations, the QDSVAR confidently gets the wrong answer: it rejects the hypothesis that the data were generated by the model. For

person is bounded, and therefore, the stochastic process for hours per person cannot literally have a unit root. Hence, according to the critique, the DSVAR procedure is misspecified with respect to all economic models and, thus, is useless for distinguishing among broad classes of models. 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 (2005a) for an eloquent defense of this position. See also Marcet (2005) for a defense of differencing in VARs.

[3] We also conduct a variety of standard lag-length tests and find that these tests do not detect the need for more lags.

the LSVAR, when demand shocks play a substantial role, the difference between the impulse response means is also large, but the confidence bands are so large that the procedure cannot distinguish between models of interest—say, between sticky price models and real business cycle models. These findings show that in practice the main claim of the SVAR literature is incorrect: the accuracy of the SVAR procedure does depend critically on the details of shocks other than technology shocks.

Our findings thus suggest that the common SVAR approach with long-run restrictions is not likely to be useful in guiding the development of business cycle theory unless demand shocks account for a trivial fraction of the fluctuations in output. We ask whether data and the literature point decisively toward an insubstantial role for demand shocks. The answer seems to be no.

We present five types of evidence which lead to that answer:

- *The central result of the SVAR literature.* For our business cycle model to generate the SVAR finding that technology shocks lead to a fall in hours, technology shocks must account for only a modest fraction of output variability, not most of it.
- *The SVAR literature itself.* The SVAR literature has argued that technology shocks account for only a modest fraction of output variability.
- *The actual observed variability in hours worked.* Our business cycle model can generate the observed variability in the U.S. hours worked series only if technology shocks account for a modest fraction of output variability.
- *The results of maximum likelihood estimation.* Based on the method of maximum likelihood estimation, differing specifications of the model and of observables indicate a sizeable range for the contribution of technology shocks. Most of the maximum likelihood estimates point to substantial errors for the impulse responses associated with both the QDSVAR and the LSVAR.
- *The growth model literature.* Studies which use the growth model to analyze business cycles contain a wide range of estimates for the contribution of technology shocks—-from zero to 100%—-with no consensus on any value in between.

We briefly examine what our findings suggest about the usefulness in practice of SVARs that use the common approach and long-run restrictions. The DSVAR literature has argued that in the data, technology shocks drive down hours on impact. We argue that this finding is highly suspect. In contrast to the DSVAR literature, the LSVAR literature in practice has been unable to guide theory because the impulse responses range so widely across studies (Christiano, Eichenbaum, and Vigfusson 2003; Francis and Ramey 2005b; Galí and Rabanal 2005). We demonstrate that some of the sharply contrasting results are driven almost entirely by small differences in the underlying data and that the responses are not stable across subsamples.

Overall, our critique challenges the dramatic recent result from the SVAR literature, which implies the death of the real business cycle model. The common SVAR approach with long-run restrictions is not a useful tool for making such judgments. The root of the problem is that the procedure compares two very different sets of statistics: empirical and theoretical impulse responses. As statistics of the data, empirical impulse responses are entirely unobjectionable. The comparison between the two sets of statistics is inappropriate because it is prone to various pitfalls, especially lag-truncation bias.

Not all SVAR procedures make such inappropriate comparisons. A preferable alternative to the common procedure is one that compares empirical impulse responses based on the data to impulse responses from identical structural VARs run on data from the model of the same length as the actual data. We call this the *Sims–Cogley–Nason approach* because it has been advocated by Sims (1989) and successfully applied by Cogley and Nason (1995). On purely logical grounds, the Sims–Cogley–Nason approach is superior to the approach we scrutinize here; it treats the data from the U.S. economy and the model economy symmetrically, thereby avoiding the problems of the common approach. Whether this alternative approach can be broadly useful has not yet been determined, but compared to the common approach, it is certainly more promising.

Our critique builds on those in studies that we discuss below, especially Sims (1971, 1972), Hansen and Sargent (1980, 1991), and Cooley and Dwyer (1998).

1. Tools for Testing

Let's start our critique of the common SVAR approach with long-run restrictions by briefly describing the two basic tools needed to apply our natural economic test: a structural VAR procedure and a business cycle model.

A. A Structural VAR Procedure

The VAR procedure we will be evaluating is a version of Blanchard and Quah's 1989 procedure used recently by Galí (1999), Francis and Ramey (2005a), and Galí and Rabanal (2005).

The procedure starts with a VAR of the form

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

where Y_t is a list (or vector) of observed variables, the B's are the VAR coefficients, and the error terms v_t have a nonsingular covariance matrix $Ev_t v_t' = \Omega$ and are orthogonal at all leads and lags, so that $Ev_t v_s' = 0$ for $s < t$. The vector Y_t is given by $(y_{1t}, y_{2t})'$, where $y_{1t} = \Delta \log(y_t/l_t)$ is the first difference of the log of labor productivity, $y_{2t} = \log l_t - \alpha \log l_{t-1}$, and l_t is a measure of the labor input. We consider two specifications of this VAR: in the *differenced* specification (DSVAR), $\alpha = 1$, so y_{2t} is the first difference in the log of the labor input; in the *level* specification (LSVAR), $\alpha = 0$, so y_{2t} is simply the log of the labor input.

This VAR, as it stands, can be thought of as a reduced form of an economic model. Specifically, the reduced-form error terms v_t have no structural interpretation. Inverting the VAR is convenient in order to express it in its equivalent moving-average form:

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

where the moving-average coefficients are defined as

$$(3) \quad C_0 = I, \quad C_1 = B_1, \quad C_2 = B_1 C_1 + B_2, \quad C_3 = B_1 C_2 + B_2 C_1 + B_3,$$

and so on. Note for later use that the sum of the moving-average coefficients $\bar{C} = \sum_{i=0}^{\infty} C_i$ is related to the VAR coefficients by

$$(4) \quad \bar{C} = \left[I - \sum_{i=1}^{p} B_i \right]^{-1}.$$

The idea behind the SVAR procedure is to use the reduced-form model (2), together with the bare minimum of economic theory, to back out structural shocks and the responses to those shocks. To see how that is done, 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, commonly referred to as a *technology shock* and a *demand shock,* respectively.

The structural model is given by

$$(5) \quad Y_t = A_0\varepsilon_t + A_1\varepsilon_{t-1} + A_2\varepsilon_{t-2} + \ldots,$$

where the A's are the structural coefficients and the $\varepsilon_t = (\varepsilon_t^z, \varepsilon_t^d)'$ represent the structural 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 Y_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. Since the technology shock is the first element of ε_t, the impulse responses to a technology shock depend only on the first column of the matrices A_i for $i = 0, 1, \ldots$.

In order for the stochastic processes for Y_t represented by (1) and (5) to coincide, we must assume that

$$(6) \quad A(L)^{-1} \text{ exists and is equal to } I - \sum_{i=1}^{p} B_i L^i,$$

where $A(L) = A_0 + A_1L + \ .\ .\ .$ and where L is the lag operator. This assumption, which we call the *auxiliary assumption*, is typically not emphasized in the literature. Under this assumption, 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 structural parameters A_i and Σ are then related to the reduced-form parameters C_i and Ω by

$$(7) \quad A_0\Sigma A_0' = \Omega \text{ and } A_i = C_iA_0 \text{ for } i \geq 1.$$

In order to identify the structural parameters from the reduced-form parameters, some other assumptions are needed. The SVAR procedure we are testing uses two identifying assumptions and a sign restriction.

One assumption is that technology shocks and demand shocks are *orthogonal.* If we interpret the structural shocks as having been scaled by their standard deviations, then we can express this assumption as $\Sigma = I$, so that $E\varepsilon_t\varepsilon_t' = I$, or equivalently as

$$(8) \quad A_0 A_0' = \Omega.$$

The other identifying assumption is a *long-run restriction*, the assumption that $\sum_{i=0}^{\infty} A_i(1,2) = 0$ and $\sum_{i=0}^{\infty} A_i(1,1) \neq 0$, where $A_i(j,k)$ is the element in the jth row and the kth column of the matrix A_i. This assumption captures the idea that demand shocks do not affect the level of labor productivity in the very long run, but technology shocks do.

To see that these assumptions identify the shocks up to a sign restriction, note that since the covariance matrix Ω is symmetric, equation (8) gives three (nonlinear) equations in the four elements of A_0. 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, or that

$$(9) \quad \bar{C}(1,1)A_0(1,2) + \bar{C}(1,2)A_0(2,2) = 0.$$

This restriction gives a fourth nontrivial equation if and only if at least one of $\bar{C}(1,1)$ or $\bar{C}(1,2)$ is nonzero, a sufficient condition for which is that a technology shock has a nonzero effect on the long-run level of labor productivity, so that

$$(10) \quad \bar{C}(1,1)A_0(1,1) + \bar{C}(1,2)A_0(2,1) \neq 0.$$

The four equations can then be solved, up to a sign convention, for the four unknown elements of A_0.

The sign restriction we will use is that a technology shock is called *positive* if it raises the level of labor productivity in the long run.[4] That is, the $(1,1)$ element of $\bar{C} A_0$ is positive, so that

$$(11) \quad \bar{C}(1,1)A_0(1,1) + \bar{C}(1,2)A_0(2,1) > 0.$$

The impulse responses for a technology shock are invariant to the sign with respect to the

[4] In some of the VAR literature, sign restrictions are viewed as convenient normalizations with no economic content. Our sign restriction, in contrast, is a restriction implied by a large class of economic models, including the business cycle models considered below. It is similar in spirit to the long-run restriction. Both restrictions use the idea that while economic models may have very different implications for short-run dynamics, they often have very similar implications for long-run behavior.

demand shock. Thus, since we focus exclusively on the impulse responses to a technology shock, for our results, the sign restriction for the demand shock is irrelevant. With these assumptions, then, we can identify the first column of each matrix A_i for $i \geq 0$, which records the impulse responses of the two variables to a technology shock. (See Appendix A for details.)

Our analysis of the problems with the common approach rests crucially on an analysis of the auxiliary assumption (6). In all of our versions of the baseline business cycle model, the auxiliary assumption is satisfied for an infinite number of lags ($p = \infty$). In practice, however, with existing data lengths, researchers are forced to run VARs with a much smaller number of lags, typically four. This lag truncation introduces a bias into the impulse responses computed using the common approach. The point of our analysis is to quantify how this lag-truncation bias varies with parameters. We also point out special circumstances under which, even though the VAR is truncated, the impulse responses to a technology shock have no lag-truncation bias.

B. A Business Cycle Model

To test the claim made for the common SVAR approach with long-run restrictions, we will use several versions of a business cycle model with multiple shocks.

The *baseline* model is a stripped-down version of business cycle models common in the literature which satisfy the two key identifying assumptions of the SVAR procedure we are evaluating, that technology and nontechnology, or *demand*, shocks are orthogonal and that demand shocks do not permanently affect the level of labor productivity while technology shocks do. The baseline model has two stochastic variables: changes in technology Z_t, which have a unit root, and an orthogonal tax on labor τ_{lt}. The model also has a constant investment tax τ_x.

Our choice of the labor tax as the demand shock is motivated by an extensive literature on business cycle models with multiple shocks. This literature grew out of the early literature on equilibrium business cycle models which focuses on models in which technology shocks account for all of the fluctuations in output. (See, for example, Kydland and Prescott 1982 and Hansen 1985.) Multiple-shock models are motivated, in part, by the inability of the early models to generate the volatility of hours observed in the data.[5] A key feature of the

[5]See, for example, Cooley and Hansen (1989); Benhabib, Rogerson, and Wright (1991); Greenwood and

multiple-shock models is that in them the fraction of variability in output due to technology shocks is much lower than in single-shock models.

A key feature of the shocks that many of these models introduce is that the shocks effectively distort consumers' labor/leisure choice. In earlier work (Chari, Kehoe, and McGrattan, forthcoming), we have shown that many of these models are equivalent to a prototype business cycle model with a labor wedge that resembles a stochastic tax on labor. We have also shown that the labor wedge and the productivity shock account for the bulk of fluctuations in U.S. data. These considerations lead us here to focus on the labor tax as a demand shock in our baseline model.

Another popular class of models includes, in addition to technology shocks, shocks that distort intertemporal margins. An investment tax mimics such distortions. In our *investment wedge* version of the business cycle model, we replace the stochastic labor tax of our baseline model with a stochastic investment tax.

In our baseline 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, where β is the discount factor and γ the growth rate of the population. Consumers maximize utility subject to the budget constraint

$$(12) \quad c_t + (1+\tau_x)[(1+\gamma)k_{t+1} - (1-\delta)k_t] = (1-\tau_{lt})w_t l_t + r_t k_t + T_t,$$

where k_t denotes the per capita capital stock, δ the depreciation rate of capital, w_t the wage rate, r_t the rental rate on capital, and T_t lump-sum taxes and where $\beta < 1$, $\gamma \geq 0$, and $0 \leq \delta \leq 1$. We assume that $U(c_t, l_t) = c_t^{1-\sigma} v(l_t)/(1-\sigma)$ in order for the model to be consistent with balanced growth.

In the model, firms have a constant returns to scale production function, $F(k_t, Z_t l_t)$, 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, where y_t denotes per capita output, is

$$(13) \quad c_t + (1+\gamma)k_{t+1} = y_t + (1-\delta)k_t.$$

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).

In our baseline model, the stochastic process for the two shocks, $\log Z_t$ and τ_{lt}, which we refer to as the *technology* and *demand shocks,* is

$$(14) \quad \log Z_{t+1} = \mu_z + \log Z_t + \log z_{t+1}$$

$$(15) \quad \tau_{lt+1} = (1-\rho_l)\bar{\tau}_l + \rho_l \tau_{lt} + \varepsilon_{lt+1},$$

where $\log z_t$ and ε_{lt} are mean zero normal random variables with standard deviations σ_z and σ_l. We let $\varepsilon_t = (\log z_t, \varepsilon_{lt})$, where these variables are independent of each other and i.i.d. over time. We refer to $\log z_t$ and ε_{lt} as the *innovations to technology and labor.* The constant $\mu_z \geq 0$ is the drift term in the random walk for technology, the parameter ρ_l is the persistence parameter for the labor tax, and $\bar{\tau}_l$ is the mean of the labor tax.

Our model satisfies the two key identifying assumptions of the SVAR approach using long-run restrictions. By construction, the two types of shocks are orthogonal. And in the model's steady state, the level of labor productivity is not affected by labor tax rates but is affected by technology levels. Thus, regardless of the persistence of the stochastic process on labor taxes, a shock to labor taxes has no effect on labor productivity in the long run.

The log-linearized decision rules are of the form

$$(16) \quad \log l_t = a(\log \hat{k}_t - \log z_t) + b\tau_{lt}$$

$$(17) \quad \log \hat{y}_t = \theta(\log \hat{k}_t - \log z_t) + (1-\theta)\log l_t$$

$$(18) \quad \log \hat{k}_{t+1} = \gamma_k(\log \hat{k}_t - \log z_t) + \gamma_l \tau_{lt},$$

where $\hat{k}_t = k_t/Z_{t-1}$, $\hat{y}_t = y_t/Z_t$, $z_t = Z_t/Z_{t-1}$, and θ is the steady-state capital share $F_k k/y$ and where here and throughout we omit constants. Note that the parameter a will be negative in our model.

The *state* of the economy in period t is $X_t = (\log \hat{k}_t, \tau_{lt-1})$. The equations governing the state variables are

$$(19) \quad \log \hat{k}_{t+1} = \gamma_k \log \hat{k}_t + \gamma_l \rho_l \tau_{lt-1} - \gamma_k \log z_t + \gamma_l \varepsilon_{lt}$$

and (15) with the constant $(1-\rho_l)\bar{\tau}_l$ omitted. We stack these equations to give the *state* *equation,* of the form

$$(20)\quad X_{t+1} = AX_t + B\varepsilon_{mt},$$

where

$$(21)\quad A = \begin{bmatrix} \gamma_k & 0 \\ 0 & \rho_l \end{bmatrix}, \; B = \begin{bmatrix} -\gamma_k & \gamma_l \\ 0 & 1 \end{bmatrix}$$

and $\varepsilon_{mt} = (\log z_t, \varepsilon_{lt})$, and the *observer equation,* of the form

$$(22)\quad Y_t = CX_t + D\varepsilon_{mt},$$

where $y_t = \big(\Delta \log \hat{y}_t / l_t,\ (1-\alpha L)\log l_t\big)$ and

$$(23)\quad C = \begin{bmatrix} \theta(1-a)\left(1-\frac{1}{\gamma_k}\right) & \theta\left[b(1-\rho_l)+\frac{(1-a)\gamma_l}{\gamma_k}\right] \\ a\left(1-\frac{\alpha}{\gamma_k}\right) & \frac{a\alpha\gamma_l}{\gamma_k}+b(\rho_l-\alpha) \end{bmatrix}, \; D = \begin{bmatrix} -\gamma_k & -\theta b \\ -a & b \end{bmatrix}$$

and where we have used (19) to substitute out for $\log z_{t-1}$. Together, the state and observer equations constitute a *state space system.* Note that eigenvalues of A are γ_k and ρ_l, which are both less than 1, so the system is stable.

Note that since the observed variables depend on y_{t-1} and l_{t-1}, it might seem necessary for the state to include $\log \hat{k}_{t-1}$ and $\log z_{t-1}$. It is not necessary, however, to include these variables because the decision rules of a growth model with a unit root in technology have a particular structure: they depend only on the difference between $\log \hat{k}_t$ and $\log z_t$.

So far we have described one particular state space system which will be convenient in proving our first proposition. In proving our second proposition, an *alternative state space system* will be more convenient. In this alternative system, the state is $S_t = (\log \hat{k}_t, \log z_t, \tau_{lt}, \tau_{lt-1})$. The *alternative state equation*, of the form

$$(24)\quad S_{t+1} = \hat{A}S_t + \hat{B}\hat{\varepsilon}_{mt+1},$$

has

$$\hat{A} = \begin{bmatrix} \gamma_k & -\gamma_k & \gamma_l & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_l & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} \gamma_k & -\gamma_k & \gamma_l & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $\hat{\varepsilon}_{mt} = (0, \varepsilon'_{mt}, 0)$. The *alternative observer equation*

(25) $Y_t = \hat{C} S_t$

has

$$\hat{C} = \begin{bmatrix} \theta(1-a)\left(1 - \frac{1}{\gamma_k}\right) & 1 - \theta(1-a) & -\theta b & \theta[b + \frac{(1-a)\gamma_l}{\gamma_k}] \\ a\left(1 - \frac{\alpha}{\gamma_k}\right) & -\frac{a}{1-\theta} & -\frac{a}{1-\theta} & \frac{a\alpha\gamma_l}{\gamma_k} - b\alpha \end{bmatrix}.$$

In making our model quantitative, we use functional forms and parameter values familiar from the business cycle literature, and we assume that the time period is one quarter. We assume that the utility function has the form $U(c,l) = \log c + \phi \log(1-l)$ and the production function, the form $F(k,l) = k^{\theta} l^{1-\theta}$. We choose the time allocation parameter $\phi = 1.6$ and the capital share $\theta = .33$. We choose the depreciation rate, the discount factor, and the growth rates so that, on an annualized basis, depreciation is 6%, the rate of time preference 2%, the population growth rate 1%, and the technology growth rate 2%. Finally, we set the mean tax labor tax $\bar{\tau}_l$ to .4.

The model's *impulse response of hours* to a technology shock is calculated recursively. We start at a steady state; set the technology innovations $\log z_0 = \Delta > 0$, $\log z_t = 0$ for $t \geq 1$; and set the labor innovations $\varepsilon_{lt} = 0$ for all t. Then, from (16) and (18), we see that the *impact effect,* namely, the impulse response in period 0, is $-a\Delta$, the effect in period 1 is $-\gamma_k a\Delta$, the effect in period $t \geq 2$ is $-\gamma_k^{t-1} a\Delta$, and so on.

In Figure 1, we plot the baseline business cycle model's impulse response of hours worked to a 1% positive technology shock. We see that in this model, on impact, a positive shock to technology leads to an increase in hours worked that persists for at least 60 quarters. The vertical axis measures the response to a 1% shock to total factor productivity (TFP). On impact, the hours increase is .42%, and the response's half-life is about 17 quarters.

Inspection of (16)–(18) makes it obvious that the model's impulse response is independent of the persistence parameter ρ_l and the variances of the innovations σ_z^2 and σ_l^2. The main claim of the SVAR literature is that the impulse response that it identifies will not depend on these parameters. Although this claim is true with an infinitely long data set, we now show that it is not true for data sets of the length of postwar data.

2. The Natural Economic Test

We test the claim of the common SVAR approach with long-run restrictions by comparing the business cycle model's impulse responses (seen in Figure 1) to those obtained by applying the SVAR procedure to data from that model, the *SVAR impulse responses*.[6] Proponents of this procedure claim that it can confidently distinguish between promising and unpromising classes of models without the researchers having to specify the details of demand shocks. We show that this claim is false by showing that the SVAR impulse responses with a finite number of lags—the number available in the small amount of actual data available—depend importantly on the parameters governing the stochastic process for demand shocks.

A. An Inessential Technical Issue

Before describing our test, we dispense with a technical issue. The common SVAR approach assumes that an autoregressive representation of the variables $\left(\Delta \log(y_t/l_t),\, l_t - \alpha l_{t-1}\right)$ exists for the models to be evaluated, in the sense that the auxiliary assumption is satisfied for some, possibly infinite, number of lags p. For the LSVAR specification ($\alpha = 0$), as we will see, the variables have an autoregressive representation. The DSVAR specification ($\alpha = 1$), however, overdifferences hours and introduces a root of 1 in the moving-average representation, which is at the edge of the noninvertibility region of roots. Hence, no autoregressive representation for the DSVAR exists. (See, for example, Fernández-Villaverde, Rubio-Ramírez, and Sargent 2005.)

This technical issue is not essential to our findings. We demonstrate that by considering, instead, a QDSVAR specification with α close to 1. We show later that as long as α is

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

less than 1, these variables have an autoregressive representation. When α is close to 1, the impulse responses of the QDSVAR and the DSVAR are so close as to be indistinguishable. In our quantitative analyses, we will set the quasi-differencing parameter α 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 Fernández-Villaverde, Rubio-Ramírez, and Sargent 2005.)

With the QDSVAR specification and an infinitely long data series, the SVAR recovers the model's impulse response. Hence, there is no issue of misspecification with the QDSVAR. (Of course, there is also no issue of misspecification with the LSVAR.)

B. Evaluation of the SVAR Claim

In our evaluation, we treat the business cycle model as the data-generating process and draw from it 1,000 data sequences of roughly the same length as our postwar U.S. data, which is 180 quarters. We run the SVAR procedure for each of the two specifications on each sequence of model data and report on the SVAR impulse responses of hours worked to technology shocks. We repeat this procedure for a wide range of parameter values for the stochastic processes and find that basically the SVAR procedure cannot do what is claimed for it.

We study the impulse response of hours worked to a technology shock and focus mainly on a simple statistic designed to capture the difference between the impulse responses of the business cycle model and the SVARs. That statistic is the *impact error,* defined as the percentage difference between the mean across sequences of the SVAR impact coefficient and the model's impact coefficient.

In Figure 2, we plot the impact errors of the QDSVAR and LSVAR specifications against a measure of the relative variability of the two shocks: the ratio of the innovation variance of the demand shock to that of the technology shock (σ_l^2/σ_z^2) for four values of the serial correlation of the demand shock ρ_l. If the SVAR claim is correct, then the errors should not vary across this measure. But they do. Notice that the impact errors for the QDSVAR specification are all negative, whereas those for the LSVAR specification are all positive. Note that an error of -100% implies that the SVAR impact coefficient is zero (instead of .42), whereas any error more negative than -100% implies that the SVAR impact coefficient is negative. The figure reveals that when the innovation variance ratio is small, so that the variance of demand shocks is small relative to that of technology shocks, the impact error is small in both specifications. As the relative variance of demand shocks increases, the absolute value of the impact error increases.

This figure contradicts the claim of the SVAR literature, that in practice the procedure accurately identifies the effect of a technology shock without having to specify the details of other orthogonal shock processes. Here, that claim translates into the claim that, in practice, the measured effect of a technology shock does not depend on the ratio of the innovation variances (σ_l^2/σ_z^2) or on the serial correlation of the demand shock ρ_l. That is clearly not correct.

In particular, Figure 2 shows that the SVAR impulse responses are quite different from those of the model when the relative variance of the demand shock is high. To better interpret Figure 2, we replace the relative variance of the demand shock by a related and more familiar statistic: the fraction of output variability due to a technology shock. We compute this fraction as the ratio of the variance of HP-filtered output with the technology shock alone relative to the variance of HP-filtered output with both shocks. We compute these variances from simulations of length 100,000. In Figure 3A, for the QDSVAR, we plot the impact error against the fraction of output variance due to a technology shock for $\rho_l = .95$ as well as the mean of the bootstrapped confidence bands across the same 1,000 sequences. Figure 3B is the analog of Figure 3A for the LSVAR.

These figures also support our main finding: the claim of the SVAR literature that this approach can confidently distinguish among models regardless of the details of the other shocks is incorrect. For the QDSVAR (Figure 3A), we see that except when the technology shock accounts for more than 80% of the variability of output, the QDSVAR confidently gets the wrong answer on impact, in the sense that the confidence bands do not include zero percent error. Moreover, unless technology shocks account for the bulk of output variability, say, more than 70%, the mean impact coefficient is negative, since the impact error is more negative than -100%.

For the LSVAR (Figure 3B), we see that except when the technology shock accounts for virtually all of the variability of output, the confidence bands in the LSVAR are so wide that this procedure cannot distinguish between most models of interest. Here, unless the technology shock accounts for much more than 90% of the variability of output, the confidence bands include negative values for the impact coefficient (that is, values for which the impact error is below −100%). Hence, as long as technology shocks account for less than 90% of output fluctuations, the LSVAR cannot distinguish between a class of models that predict a negative impact (like sticky price models) and a class of models that predict a positive impact (like real business cycle models). In terms of the impact error, note that when technology shocks account for less than 45% of the variability of output, the mean impact error is greater than 100%. Note also that the confidence bands for the LSVAR are wider than those for the QDSVAR.

Clearly, for neither specification is the claim of the SVAR literature supported by our test.

C. Statistical Tests for a Particular Parameter Set

So far we have focused on the means of the impact error and the means of the associated confidence bands across simulations. Here we ask whether a researcher can accurately detect whether the data are generated by our business cycle model or by some other model. Since providing these details for a wide range of parameters is cumbersome, we focus on a particular parameter set which is linked to the work of Galí (1999).

The key parameter is the measure we have used above, the relative variability of technology to demand shocks. The SVAR literature together with our business cycle model can also be used to indirectly infer this parameter. The central finding of the SVAR literature based on long-run restrictions is Galí's (1999) widely noted finding that a positive technology shock drives down hours worked on impact. (Indeed, this finding is the genesis of the recent upsurge in interest in this branch of the SVAR literature.) In evaluating the SVAR procedure, we think that if the procedure is a good one, then when it is applied to data generated from our model, it should be able to reproduce Galí's central finding. We therefore investigate what the ratio of the innovation variances must be in order for the mean of the impact coefficient of hours to a technology shock obtained from the QDSVAR to be similar to Galí's (1999) impact coefficient.[7]

In Figure 4A, we plot some results based on Galí's parameters. In the left graph, we show the histogram of the QDSVAR's impact coefficient over the 1,000 sequences. The histogram shows that almost all of these coefficients are negative. The right graph of Figure

[7] We do not attempt to perform a similar exercise with respect to the LSVAR literature because, as we document below, the impact coefficients range widely across studies, from large positive numbers to large negative ones.

4A reports the range of estimated impulse responses over these 1,000 sequences for 12 quarters after the shock as well as the business cycle model's impulse response. We construct the 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%. The figure shows that the impulse responses for essentially all of the QDSVAR simulations are quite different from those of the business cycle model. The SVAR ranges do not, in fact, include the model's response.

Now, for each of the 1,000 sequences, we suppose that a researcher tests 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 do not come from our business cycle model about 88% of the time. Figure 4B displays the mean impulse response across these 1,000 sequences and the mean of the bootstrapped confidence bands across the same sequences. This figure gives some intuition for why a researcher would typically draw the wrong inference.

Figures 5A and 5B, the analogs of Figures 4A and 4B for the LSVAR specification, provide some intuition for our result that the LSVAR is not useful in distinguishing among many classes of models. From the histogram in the left graph of Figure 5A, we see that the range of impact coefficients is very wide. For example, in the right graph of Figure 5A we see that 95% of the impact coefficients lie between $-.60$ and 1.68.

For the LSVAR as for the QDSVAR, we now suppose that for each of the 1,000 sequences, a researcher tests the hypothesis that the SVAR's impact coefficient equals 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, what if the researcher tests the hypothesis that the impact coefficient of the LSVAR equals zero at the 5% significance level? Such a researcher would essentially never reject this hypothesis either.

These findings, together with the other graphs of Figures 5A and 5B, suggest that with data of the same length as postwar U.S. data, the LSVAR cannot differentiate between models with starkly different impulse response functions, for example, between 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, Galí, López-Salido, and Vallés (2003) construct a simple sticky price model in which the monetary authority follows a Taylor rule; using this model, they show that hours rise in response to a technology shock. They also show that if monetary policy is not at all accommodative, then 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, as the right panel of Figure 5B shows, and within the 95% confidence bands, as Figure 5B shows.

So far we have simply assumed that researchers must choose either the QDSVAR specification or the LSVAR specification for all samples. In practice, researchers often conduct tests to determine which specification is preferable for their particular samples. Typically, they conduct unit root tests to determine whether in the VAR hours should be specified in levels or in first differences. Here we ask whether our findings are robust to a procedure which mimics the procedures conducted in practice. They are. We focus here on the Galí parameters because at these values the model reproduces the central finding of the SVAR literature. We experimented with other parameter values and got similar results.

We first consider unit root tests. For each of the 1,000 sequences generated from our model, we conducted an augmented Dickey-Fuller unit root test on hours (with a trend and four lags). We find that the test does not reject a unit root for most of the sequences. For example, with $\rho_l = .95$, it does not reject a unit root in about 85% of the sequences. We get similar results from other unit root tests.

We also experiment with variants of the SVAR procedure. For the QDSVAR specification, we retained only sequences which passed the unit root test. Our findings are virtually identical to those we have reported. For the LSVAR specification, we retained only sequences which failed the unit root test. Here also our results are virtually identical to those we have reported.

Researchers often conduct lag-length tests to determine the appropriate number of lags. In an attempt to mimic a variant of the common approach which uses both lag-length tests and unit root tests, we experiment with variants of the SVAR procedure. For the QDSVAR specification, we retained only sequences which passed both the unit root test and the standard lag-length tests (described in more detail below). We also allowed the lag length for each sequence to be determined by the lag-length tests. Again, our findings are virtually identical to those reported above. For the LSVAR specification, we retained only sequences which passed the lag-length test, and we allowed the lag length for each sequence to be determined by the lag-length tests. Here also our results are virtually identical to those we have reported.

Considering the results from all our quantitative analysis, we conclude that for both specifications, the claim of the SVAR literature is not correct.

3. Analyzing the SVAR's Impulse Response Error

Here we investigate why the SVAR procedure fails our test. We determine that the problem with the procedure rests crucially on the auxiliary assumption (6), that Y_t has an autoregressive representation well-approximated with a small number of lags. The impact error is large in our test when the business cycle model does not satisfy this assumption and small when it does. In all of the versions of our business cycle model, the auxiliary assumption is satisfied with an indefinite number of lags ($p = \infty$). In practice, however, researchers are forced by the existing data lengths to run SVARs with a small number of lags, typically four. This lag truncation introduces a bias into the SVAR impulse responses. We here quantify how the lag-truncation bias varies with parameters and point out special circumstances under which, even though the VAR is truncated, the impulse responses to a technology shock have no such bias. These special circumstances include the case in which the nontechnology plays a trivial role and when capital plays a trivial role.

A. Analysis of the Auxiliary Assumption

Here we analyze the SVAR's auxiliary assumption for general state space systems and draw out its implications for our two-variable system. We prove two propositions which provide intuition for when the SVAR procedure performs poorly and when it performs well. First we prove that when the number of observed variables is the same as the number of (nontrivial) shocks, the associated VAR satisfies the auxiliary assumption with the number of lags $p = \infty$. Then we prove that when the alternative state is an invertible function of the observed variables, the observed variables have a first-order VAR representation (with a singular covariance matrix for the shocks). For our two-variable VAR, the first proposition implies that when the variance of the demand shock is positive, the VAR has $p = \infty$, so that the auxiliary assumption fails with $p = 4$. The second proposition implies that when the variance of the demand shock is zero, the VAR has $p = 1$, so that the auxiliary assumption is satisfied with $p = 4$. Together these propositions demonstrate that the small number of lags is at the heart of the SVAR problem when the demand shocks play a nontrivial role and demonstrate why the SVAR procedure works well when demand shocks play a trivial role.

Same Number of Variables as Shocks

Consider a state space system of the form (20) and (22) for general matrices A, B, C, and D. Standard arguments (as in Fernández-Villaverde, Rubio-Ramírez, and Sargent 2005) lead to the following result for any state space system with the same number of observable variables as shocks:

PROPOSITION 1. (*Existence of an Infinite-Order Autoregressive Representation*) Consider any state space system of the form (20) and (22), and assume the system has the same number of observables as shocks, so that the matrix D is square. Suppose that D is invertible, the eigenvalues of A are less than 1, and the eigenvalues of $A-BD^{-1}C$ are strictly less than 1. Then the model's moving-average representation is invertible and the model's autoregressive representation of Y_t is given by

$$(26)\quad Y_t = B_{m1}Y_{t-1} + MB_{m1}Y_{t-2} + M^2B_{m1}Y_{t-3} + \ldots + D\varepsilon_{mt},$$

where the decay matrix M is given by $M = C[A - BD^{-1}C]C^{-1}$.

Proof. Since the matrix D is invertible, $\varepsilon_{mt} = D^{-1}(Y_t - CX_t)$. Substituting into the state equation and rearranging gives $[I-(A-BD^{-1}C)L]X_{t+1} = BD^{-1}Y_t$, where L is the lag operator. If the eigenvalues of $A - BD^{-1}C$ are strictly less than 1 in modulus, then we can write $X_{t+1} = \sum_{j=0}^{\infty}[A - BD^{-1}C]^j BD^{-1}Y_{t-j}$. Using this equation to substitute for X_t in the observer equation gives the desired autoregressive representation:

$$(27)\quad Y_t = C\sum_{j=0}^{\infty}[A - BD^{-1}C]^j BD^{-1}Y_{t-j-1} + D\varepsilon_{mt}.$$

We can rewrite this representation as (26). Note that $B_{m1} = CBD^{-1}$ and that $B_{m2} = C[A - BD^{-1}C]BD^{-1}$, so that $B_{m2} = MB_{m1}$, where $M = C[A - BD^{-1}C]C^{-1}$. Likewise, $B_{mj+1} = MB_{mj}$ for all j.

Note that if the roots of A are less than 1 in modulus, then the model has a moving-average representation in terms of past values of the economic shocks ε_{mt} of the form

$$(28)\quad Y_t = D\varepsilon_{mt} + CB\varepsilon_{mt-1} + CAB\varepsilon_{mt-2} + CA^2B\varepsilon_{mt-3} + \ldots.$$

Since (27) and (28) are representations of the same stochastic process, the moving-average representation is invertible if the roots of both A and $A - BD^{-1}C$ are strictly less than 1 in modulus. *Q.E.D.*

Next we show that for a wide range of parameters, the sufficient conditions in Proposition 1 are satisfied in our model in which the matrices in the state space system are specified by (21) and (23). The eigenvalues of A are ρ_l and γ_k. We have assumed that ρ_l is less than 1, and it is easy to show that γ_k is too. Straightforward but tedious computations yield that the eigenvalues of $A - BD^{-1}C$ are α and $(\gamma_k - \gamma_l a/b - \theta)/(1 - \theta)$. We then have the following corollary:

Corollary 1. (*Our Model's Autoregressive Representation*) The eigenvalues of $A - BD^{-1}C$ are less than 1 if $\alpha \in [0, 1)$ and $\gamma_k - \gamma_l a/b < 1$.

For a wide range of parameters for our business cycle model, D is invertible and $\gamma_k - \gamma_l a/b < 1$. Thus, for a wide range of parameters, our model satisfies the sufficient conditions of Proposition 1 and, hence, satisfies the auxiliary assumption with $p = \infty$. Since our model also satisfies the two key identifying assumptions of the SVAR procedures, we have that if a VAR with an infinite number of lags were run on an infinitely long sample of data generated by our model, then the impulse responses from both the QDSVAR specification and the LSVAR specification would coincide exactly (in the relevant sense of convergence) with those of the model. We emphasize that our model does not suffer from the invertibility problems discussed by Hansen and Sargent (1980) and Fernández-Villaverde, Rubio-Ramírez, and Sargent (2005). Moreover, neither specification suffers from issues of identification, overdifferencing, or specification error. Without more detailed quantitative analyses, theory provides no guidance as to which specification is preferable.

Note that standard linear algebra results imply that the eigenvalues of $A - BD^{-1}C$ equal those of the decay matrix M. Given our model parameters, we have that for the QDSVAR specification (including the quasi-differencing parameter $\alpha = .99$), the eigenvalues for M are $\lambda_1 = .99$ and $\lambda_2 = .96$, whereas for the LSVAR specification, they are $\lambda_1 = 0$ and $\lambda_2 = .96$. At our model parameters, for both specifications, the largest eigenvalue is close to 1. 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. It is not surprising, then, that the SVAR procedure performs poorly when both shocks have nontrivial variances.

More Variables Than Shocks

Now consider situations with more observed variables than shocks.

We first develop sufficient conditions for the existence of a first-order autoregressive representation for the observed variables. The basic idea is that the observed variables have such a representation if the underlying state can be uncovered from them.

In developing sufficient conditions for this to be true, it is convenient to work with the alternative state space system (described in the business cycle model section) of the form

$$(29)\quad S_{t+1} = \hat{A}S_t + \hat{B}\hat{\varepsilon}_{mt+1} \text{ and } Y_t = \hat{C}S_t,$$

where the eigenvalues of $\hat{A}$ are less than 1 in modulus, so that the system is stable.

PROPOSITION 2. (*Existence of a First-Order Autoregressive Representation*) In a state space system of the form (29) with $\hat{C}$ invertible, the observed variables Y_t have an AR1 representation.

Proof. Substituting $S_t = \hat{C}^{-1}Y_t$ into the state equation and premultiplying by $\hat{C}$ gives

$$Y_{t+1} = \hat{C}\hat{A}\hat{C}^{-1}Y_t + \hat{C}\hat{B}\hat{\varepsilon}_{mt+1},$$

so that the VAR associated with the state system has only one lag. *Q.E.D.*

In most business cycle models, for this proposition to apply, the number of observed variables must be greater than the number of shocks. To see why, note that the state space representation of business cycle models typically must include an endogenous state variable like capital in addition to the exogenous shocks. Thus, if the number of observed variables equals the number of shocks, then the dimension of the state S_t is greater than the dimension of the observed variables Y_t, so that the observer matrix $\hat{C}$ is not invertible and Proposition 2 cannot apply. For example, in a system with one endogenous state variable, a necessary condition for Proposition 2 to apply is that the number of observed variables be at least one more than the number of shocks. (The sufficient conditions, of course, are stronger.)

We can apply Proposition 2 to our model with two observed variables if the variance of the demand shock is zero. In the alternative state space system, the state is then $S_t =$

$(\log \hat{k}_t, \log z_t)$. The matrices $\hat{A}$ and $\hat{B}$ are given in the alternative state equation

$$\begin{bmatrix} \log \hat{k}_{t+1} \\ \log z_{t+1} \end{bmatrix} = \begin{bmatrix} \gamma_k & -\gamma_k \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \log \hat{k}_t \\ \log z_t \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \log z_{t+1} \end{bmatrix}$$

and the matrix $\hat{C}$ in the alternative observer equation

$$\begin{bmatrix} \Delta(\log y_t/l_t) \\ \log l_t - \alpha \log l_{t-1} \end{bmatrix} = \begin{bmatrix} \theta(1-a)\left(1-\frac{1}{\gamma_k}\right) & 1-\theta(1-a) \\ a\left(1-\frac{\alpha}{\gamma_k}\right) & -\frac{a}{1-\theta} \end{bmatrix} \begin{bmatrix} \log \hat{k}_t \\ \log z_t \end{bmatrix}.$$

As long as $\hat{C}$ is invertible, the observables have a first-order autoregressive representation. Along with the fact that impulse responses are continuous in the parameters, Proposition 2 provides some intuition for why in our model, when the variance of the demand shock decreases to zero, the VAR on the observed variables is increasingly well-approximated by a VAR with one lag.

To see that having more variables than shocks is not a sufficient condition for Proposition 2 to apply, suppose that in our model with two variables, the variance of the technology shock is zero. In the alternative state space system, the state is then $S_t = (\log \hat{k}_t, \tau_{lt}, \tau_{lt-1})$. Hence, the dimension of the alternative state is greater than the dimension of the observed variables, and the state cannot be uncovered from the observed variables. Thus, Proposition 2 does not apply.

Note that a version of Proposition 2 does apply if the dimension of the observed variables exceeds the dimension of states. In such a case, we can augment the state with dummy variables and then apply Proposition 2.

Proposition 2 sheds light on a literature that argues that sometimes SVARs with long-run restrictions work well. For example, Fernández-Villaverde, Rubio-Ramírez, and Sargent (2005) show that in Fisher's (2006) model, the population estimates from an SVAR procedure with one lag closely approximate the model's impulse responses. In Appendix B we show that Fisher's VAR system has enough observed variables so that it is a special case of Proposition 2.

B. Decomposition of the Impact Error: Two Biases

>From the discussion following Proposition 1, we know that if a VAR with an infinite number of lags were to be estimated on an infinite amount of data, then the impulse responses from the common approach would converge, in the usual sense, to the theoretical impulse responses. This discussion implies a natural decomposition of the impact error into that due to small-sample bias and that due to lag-truncation bias. We do this decomposition here and find that the SVAR error is primarily due to the lag-truncation bias.

Let $\bar{A}_0(p, T)$ denote the mean of the small-sample distribution of the SVAR impulse response when the VAR has p lags and the length of the sample is T. In practice, this mean is approximated as the mean across a large number of simulations. Note that the above discussion implies that $\bar{A}_0(p = \infty, T = \infty)$ coincides with the model's theoretical impulse response. That convergence implies that the (level of the) impact error associated with our implementation of the common approach is

$$\bar{A}_0(p = 4, T = 180) - \bar{A}_0(p = \infty, T = \infty).$$

We can decompose this error into two parts:

$$\left[\bar{A}_0(p = 4, T = 180) - \bar{A}_0(p = 4, T = \infty)\right] + \left[\bar{A}_0(p = 4, T = \infty) - \bar{A}_0(p = \infty, T = \infty)\right].$$

The term in the first brackets is the *small-sample bias,* the difference between the mean of the SVAR impulse response over simulations of length 180 when the VAR has four lags and the SVAR population impulse response when the VAR has four lags. The term in the second brackets is the *lag-truncation bias,* the difference between the SVAR population impulse response when the VAR has four lags and the model's theoretical impulse response.

That VARs have small-sample biases has been known at least since Hurwicz (1950): even when the true model has a VAR with four lags, the estimated coefficients are biased in small samples.

This type of bias is small for our model: for the QDSVAR specification, it is very small, and for the LSVAR specification, it is small compared to the lag-truncation bias. These findings can be seen in Figures 6A and 6B which display the biases for the Galí parameters. For each of the two specifications, the figures show the percentage difference between the mean $A_0(p = 4, T = 180)$ and the $A_0(p = \infty, T = \infty)$, labeled *small-sample mean,* and the percentage difference between $A_0(p = 4, T = \infty)$ and $A_0(p = \infty, T = \infty)$, labeled *population.* These, again, represent the small-sample bias and the lag-truncation bias, respectively. Note in the figures that the small-sample bias does not vary much with the relative variance of the demand shock, so that the comparative static properties of the lag-truncation bias are very similar to those of the impulse response error.

These findings lead us to focus on the lag-truncation bias. As we have proven, with a sufficiently large number of lags, the lag-truncation bias becomes arbitrarily small. We ask how many lags are needed here for the lag-truncation bias to be small with the Galí parameters. The answer, we find, is too many.

Figure 7 displays the QDSVAR responses for lag lengths p ranging from 4 to 300. Notice that even with 20 lags, the lag-truncation bias of the QDSVAR specification is large. On these graphs, note that the convergence to the model's impulse response function is not monotonic. Finally, note that more than 200 lags are needed for the lag-truncation bias of the QDSVAR to be small.

Figure 8 shows the impulse responses from the LSVAR for lag lengths p ranging from 4 to 100. Here, as with the QDSVAR, we see that the impulse response from the LSVAR is a good approximation to the model's impulse response only for an extremely large number of lags. In practice, of course, accurately estimating VARs with so many lags is not feasible.

To understand the source of the lag-truncation bias, recall that in computing the impulse responses from a VAR, we use the estimated covariance matrix Ω and the estimated sum of the moving-average coefficients matrix $\bar{C}$. In unreported work, we show that the primary source of the lag-truncation bias is that the estimated matrix $\bar{C}$ is a poor approximation to the true matrix $\bar{C}_m$ from the model.

To get some intuition for why with four lags $\bar{C}$ is a poor approximation to $\bar{C}_m$, recall from (4) that

$$(30)\quad \bar{C} = \left[I - \sum_{i=1}^{4} B_i \right]^{-1}$$

while

$$(31)\quad \bar{C}_m = \left[I - \sum_{i=1}^{\infty} B_{mi} \right]^{-1}.$$

To develop the intuition for why the estimated sum $\sum_{i=1}^{4} B_i$ is a poor approximation to the model's sum $\sum_{i=1}^{\infty} B_{mi}$, note that Proposition 1 implies that the autoregressive coefficients B_{mi} in the model decay according to the matrix M. As we have shown, the largest eigenvalue of M is close to 1, so that the estimated sum is a poor approximation to the model's sum.

C. Lag-Length Tests

We have argued that the main source of the error in the common SVAR procedure is the lag-truncation bias. Here we ask whether a researcher applying the SVAR procedure and standard methods of detecting appropriate lag lengths to data for our model would detect the business cycle model's need for more than four lags. We computed a variety of lag-length tests, including the Akaike criterion, the Schwartz criterion, and a likelihood ratio test on data generated from our model. Here we report on the results for the Galí parameters. We find that none of these tests detects the need for more lags.

We generated from our model 1,000 sequences of length 180 for the variables used in the two SVAR procedures. For the QDSVAR specification, we find that the Akaike criterion selects a lag length of four or fewer in over 98.6% of the simulations and the Schwartz criterion, in all of them. The likelihood test does not reject four lags in favor of five lags in over 92.8% of the simulations. In Figure 9A, we graph the mean of the Akaike and Schwartz criteria for the QDSVAR specification against the number of lags. The means of both of these criteria are minimized at one lag.

We repeated the lag-length tests for the LSVAR specification. Now the Akaike criterion selects a lag length of four or fewer in over 99.6% of the simulations and the Schwartz criterion, again, in all of them. The likelihood test does not reject four lags in favor of five lags in over 94.4% of the simulations. In Figure 9B, we graph the mean of the Akaike and Schwartz criteria for the LSVAR specification against the number of lags. The means of both of these criteria are again minimized at one lag.

Taken together, these results suggest that with samples of roughly the same length as U.S. data, a researcher using standard methods would not detect the need for more lags for the VAR in either specification. At a mechanical level, the reason the Akaike and Schwartz lag-length tests do not detect the need for more lags is simple. These tests balance the gain in the fit of the model from adding more parameters against a fixed penalty for doing so. As more parameters are added, the gain in the fit of the model is smaller than the penalty.

D. The Role of Capital

One reason that the lag-truncation bias is large when the number of lags is small is the presence of capital in the business cycle model. We demonstrate that by proving that when the capital share is zero, the lag-truncation bias is zero even with a one-lag SVAR of the form

(32) $Y_t = B_1 Y_{t-1} + v_t$ with $E v_t {v_t}' = \Omega$.

PROPOSITION 3. (*Zero Capital Share*) When the capital share θ is zero, the lag-truncation bias is zero for the impulse response from a technology shock in an SVAR procedure with one lag.

Proof. When the capital share $\theta = 0$, the theoretical impulse response of labor to a technology shock is identically zero and the impulse response of the change in labor productivity to a technology shock is one on impact and zero thereafter. We will show that, in expectation, the impulse response for labor and the change in labor productivity constructed from a one-lag ordinary least squares (OLS) autoregression will have this form. That is, $A_j(2,1) = 0$ for all j, $A_0(1,1) = 1$, and $A_j(1,1) = 0$ for $j \geq 1$.

The log-linearized equations for the business cycle model are now

(33) $\log l_t - \alpha \log l_{t-1} = b\tau_{lt} - \alpha b \tau_{lt-1}$

(34) $\Delta \log(y_t/l_t) = \log z_t$.

Clearly, in expectation, the coefficients of the OLS regression will have the form $B_1 = \text{diag}[0, \beta_{22}]$ for some β_{22}. The expectation of the estimated covariance matrix, Ω, will have positive elements on the diagonal and zeros off the diagonal with $\Omega(1,2) = 0$. Since we normalized the variance of the technology shock to one, (34) implies that $\Omega(1,1) = 1$.

>From (30) and the form of B_1, we have that

$$\bar{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1/(1-\beta_{22}) \end{bmatrix}.$$

The long-run restriction (9) implies that $\bar{C}(1,1)A_0(1,2) = 0$, so that $A_0(1,2) = 0$. Since

$A_0 A_0' = \Omega$, we know that

$$(35) \quad A_0(1,1)^2 + A_0(1,2)^2 = \Omega(1,1) = 1$$

$$(36) \quad A_0(1,1)A_0(2,1) + A_0(1,2)A_0(2,2) = \Omega(1,2) = 0.$$

From (35) and $A_0(1,2) = 0$ and our sign restriction, we know that $A_0(1,1) = 1$. Since $A_0(1,1) \neq 0$ and $A_0(1,2) = 0$, (36) implies that $A_0(2,1) = 0$. For the subsequent coefficients A_j, recall from (7) that $A_j = C_j A_0$. From (3), $C_j = B_1^j$, so that $C_j = \text{diag}[0, y]$ for some y. Hence, $A_j(2,1) = 0$ and $A_j(1,1) = 0$ for $j \geq 1$. *Q.E.D.*

Note that, at least when the quasi-differencing parameter α is nonzero, observed variables do not have a first-order autoregressive representation. In particular, the lag-truncation bias for a demand shock will not be zero.

We also experimented with increasing the depreciation rate as another way of reducing the importance of capital. We found that when the depreciation is so high that capital essentially depreciates completely within a year, the lag-truncation bias is close to zero.

4. Does Adding Variables and Shocks Help?

So far we have focused on an SVAR with just two variables—the log difference of labor productivity and a measure of the labor input—and two shocks—one to technology and one to demand. In the SVAR literature, researchers often check how their results change when they add one or more variables and shocks to the SVAR. Would such an alteration to the SVAR we have been testing help it with our business cycle model? We find that with additional variables and shocks, the SVAR procedure can sometimes uncover the model's impulse response to shocks, but only if the states are an invertible function of the observables.

Which variables should be added to the SVAR? How about some form of capital? Our discussion of Proposition 3 suggests that one of the problems with the SVAR specification is that it does not include such a variable. In our business cycle model, the relevant state variable is $\hat{k}_t = k_t/Z_{t-1}$. However, since Z_{t-1} is not observable, we cannot include $\hat{k}_t$ itself in the SVAR. We consider instead 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 of the capital stock $\log k_{t+1} - \log k_t$. One conjecture is that including such variables might diminish the need for estimating a large number of lags in the SVARs, so that the specifications with few lags will yield accurate measures of the model's response to a technology shock. This conjecture turns out to be, in general, incorrect.

As we show in a separate 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 decay according to the matrix M, in a manner similar to that in Proposition 1. When we add the capital/output ratio, one of the eigenvalues of M is $-\infty$, whereas 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 by Hansen and Sargent (1991), we do not investigate them here.

So now we turn to the alternative state space representation of a three-shock model and ask if we can find a third variable for which the SVAR specification mimics the model's state space representation. In the LSVAR specification, we find that if we add k_{t+1}/y_t, the ratio of the capital stock in period $t+1$ to output in period t, then the SVAR representation mimics the state space representation. In this exceptional case, the lag-truncation bias of the LSVAR procedure is zero.

This finding does not imply, however, 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 lag-truncation bias of the SVAR procedure is not zero. More generally, across models, a careful examination of the state space representation for each model could lead to a different SVAR specification for each model. If so, estimating the state space representation implied by the model directly is both safer and more transparent.

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 think that the capital stock is poorly measured. The issues of invertibility and measurement lead us to use the investment/output ratio to capture the influence of the capital-like variable.

Let's see what happens with this ratio included. Consider an SVAR with three variables and three shocks. The third variable is the log of the investment/output ratio x_t/y_t, where $x_t = (1+\gamma)k_{t+1} - (1-\delta)k_t$. Here, in addition to the growth of labor productivity and the measure of labor, Y_t includes the investment/output ratio. We let the investment tax be the third shock. We assume that taxes on investment follow the autoregressive process

$$(37) \quad \tau_{xt+1} = (1-\rho_x)\bar{\tau}_x + \rho_x \tau_{xt} + \varepsilon^x_{mt+1},$$

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. The standard deviation of ε^x_{mt} is σ_x.

For this altered SVAR, Propositions 1 and 2 immediately apply. The eigenvalues of $A - BD^{-1}C$ equal those of M and are given by α, $(1-\delta)/(1+g_y)$, and 0, where g_y is the growth rate of (total) output. The analog of Corollary 1 is

COROLLARY 2. The eigenvalues of $A - BD^{-1}C$ are less than 1 if $\alpha \in [0, 1)$.

Given our parameters, the eigenvalue $(1-\delta)/(1+g_y) = .98$. 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 and, hence, (30) is a poor approximation to (31). Interestingly, the largest eigenvalue of the decay matrix M is roughly the same in the two- and three-variable SVARs, so that adding another variable does not seem to diminish the need for many lags in the VAR.

We also have experimented with four-variable SVARs and four shocks. Relative to the baseline business cycle model, we have added shocks to the tax on investment and government consumption. In the SVAR specifications, we have added the investment/output ratio and the consumption/output ratio as variables. Proposition 1 applies to this case, and the four eigenvalues of the decay matrix are given by $\alpha, (1-\delta)/(1+g_y), 0$, and 0.

As we have noted in our discussion following Proposition 2, having more observed variables than shocks is not a recipe for success. For a three-variable SVAR, say, $Y_t = \big(\Delta \log(y_t/l_t), l_t, x_t/y_t\big)$, and only two shocks, the technology shock and the labor tax shock, the three observed variables in the QDSVAR do not have a first-order autoregressive representation. The reason is that the alternative state $S_t = (\log \hat{k}_t, \log z_t, \tau_{lt}, \tau_{lt-1})$ has four variables, so that with three observed variables, the matrix $\hat{C}$ in the observer equation is not invertible.

Next we examine a quantitative version of our three-shock model with a three-variable LSVAR with $Y_t = \big(\Delta \log(y_t/l_t), l_t, x_t/y_t\big)$ and show that the lag-truncation bias and the impact errors are large even for small variances of the investment tax shock. Figure 10 displays the lag-truncation bias (labeled *population*), the impact error (labeled *small-sample mean*), and the confidence bands for the LSVAR with four lags against the percentage of the HP-filtered output due to the investment tax shock for $\rho_x = .95$, with the Galí parameters for the labor tax shock. We get similar results for other values of ρ_x.

Figure 10 shows that the lag-truncation bias is zero when the variance of the third shock is zero, as indicated by our earlier discussion. The figure also shows that this error increases rapidly with the variability due to the investment tax shock. For example, the error is over 100% if the variance of output due to the investment tax shock is 7% or more. Interestingly, even when the variance due to the third shock is essentially zero, the impact error is positive due to the small-sample bias. Finally, the figure shows that the SVAR's confidence bands are extremely wide even when the variance due to the third shock is tiny.

One might interpret Proposition 2 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) account for the bulk of fluctuations in the data. Figure 10 shows that this interpretation should be treated with caution.

5. Is the Evidence Decisive?

As we have seen, the QDSVAR and the LSVAR specifications do reasonably well only when technology shocks account for virtually all of the variability in output. How likely is that to be true? If demand shocks are usually trivial, then SVARs may be useful after all. Here we examine five types of evidence on the relative size of technology and demand shocks. We show that this evidence is far from decisive. Four types of evidence lead to the conclusion that demand shocks must play a significant role in output's variability, and one type of evidence points to a wide range of estimates for the contribution of technology shocks.

Therefore, any claim that the data definitively imply that technology shocks account for virtually all of the variability in output is exceptionally difficult to support. The data do not rule out the possibility that demand shocks play a nontrivial role in output variability; indeed, some aspects of the data suggest that they play a substantial role. If demand shocks play a nontrivial role, then both SVAR specifications perform poorly.

In presenting this evidence, we use results both from the baseline business cycle model, in which a labor tax is the second shock, and an investment wedge model, in which the investment tax is the second shock. In the investment wedge model, we assume that taxes on investment follow the process described in (37).

A. Evidence Based on the SVAR Central Finding

One type of evidence on the relative size of the two shocks is based on the central result of the SVAR literature, Galí's (1999) widely noted finding that a positive technology shock drives down hours worked on impact. For our SVAR model to generate that finding, technology shocks must account for only a modest fraction of output variability.

We demonstrate that in Figure 11. There we plot the mean of the SVAR impact coefficients against the ratio of the innovation variance for both the QDSVAR (top graph) and the LSVAR (bottom graph) specifications, fixing ρ_l at .95. (In the technical appendix (Chari, Kehoe, and McGrattan 2005), we repeat this experiment for several values of ρ_l and find similar results.) In the top graph of Figure 11, the upper horizontal solid line (labeled "To reproduce Galí's estimate") is set so that the mean impact coefficient equals $-.33$, the impact coefficient consistent with Galí's (1999) bivariate DSVAR.[8] We thus refer to the associated parameters as the *Galí parameters*.

This figure can be used to indirectly infer the relative contribution of technology and demand shocks. As we have argued above, in evaluating SVARs with long-run restrictions, the model's parameters should be such that the SVAR on the model's data reproduces the central finding of the SVAR literature: a positive technology shock drives down hours on impact. We have investigated above what the ratio of the innovation variances must be in order for the mean of the QDSVAR's impact coefficient of hours to a technology shock to be similar to Galí's (1999) impact coefficient. As Figure 11 and Table 1 indicate, at this value of the impact coefficient, the variance of output due to a technology shock is roughly 50% for our baseline model.

Figure 11 shows that if demand shocks are unimportant relative to technology shocks, then the QDSVAR impact coefficient is positive and, therefore, of the opposite sign of that estimated by Galí. For example, if the demand shock accounts for more than 30% of the variability in output, then the QDSVAR error is greater than 100, so that the impact coeffi-

[8] Galí (1999) reports that on impact, a one standard deviation technology shock leads to a $-.38\%$ change in hours. We convert this statistic to the response to a 1% technology shock, z, by dividing his statistic by the standard deviation of the technology shock. We use Prescott's (1986) measure of the standard deviation of an innovation to total factor productivity σ_{TFP} to construct the standard deviation of the technology shock σ_z. The relationship between these standard deviations is $\sigma_z = \sigma_{TFP}/(1-\theta)$. Prescott measures σ_{TFP} to be .763, and our capital share is $\theta = .33$, so that after conversion Galí's statistic becomes $-.33$ $(= -.38/\sigma_z)$.

cient is positive. Put differently, if demand shocks were this small, then there would be no controversy over what happens after a technology shock and, hence, no SVAR literature to critique.

We also conducted a similar exercise in the investment wedge model. As Table 2 shows, with this model, when the standard deviation of investment tax shocks is set so that the QDSVAR impact coefficient mimics Gali's coefficient, technology shocks account for 46% of the variability in output and the error in the LSVAR impact coefficient is 188%.

B. Other SVAR Evidence

The SVAR literature also provides direct evidence on the modest relative contribution of technology shocks to the variability in output.

Galí and Rabanal (2005), for example, use a VAR procedure on various measures of U.S. output and employment. For their measure of the fraction of output variability due to technology shocks, these researchers use the variance of the estimated business cycle component of the historical series for output associated with technology shocks relative to the sample variance of output. For their LSVAR specification, Galí and Rabanal's point estimates range from 3% to 37%. (For the DSVAR specification, they range from 6% to 31%.) Other SVAR studies find similar ranges.

These findings suggest that, at least for the purpose of evaluating the SVAR procedure, models in which demand shocks do not account for the bulk of the fluctuations in output are not interesting. Substantively, of course, as we have seen, impact errors and confidence bands associated with the SVAR procedure are large precisely when demand shocks do account for the bulk of the fluctuations.

C. Evidence Based on the Volatility of Hours in U.S. Data

A third type of evidence is based on the volatility of the U.S. time series hours worked. We ask how large demand shocks must be if our business cycle model is to reproduce the volatility of this series. One motivation for asking this question is that many of the recent developments in business cycle theory are driven by the observation that business cycle models with only technology shocks cannot produce anywhere near the volatility of hours in the data. This failure is particularly marked when the technology shock has a unit root. We find that for our model to reproduce the actual volatility of U.S. hours, demand shocks must be so volatile that the SVAR procedure performs poorly.

For example, suppose we set the standard deviation of the technology shock, σ_z, to reproduce Prescott's (1986) measure of the standard deviation of an innovation to total factor productivity σ_{TFP}. (The relationship between these standard deviations is $\sigma_z = \sigma_{TFP}/(1-\theta)$.) If the variance of the demand shock σ_l is zero, then the volatility of per capita hours in the model is only about 5% of that of per capita hours for the U.S. economy in the data.[9]

We then ask, what must be the volatility (standard deviation) of the demand shock, σ_l, in order to reproduce the observed volatility in hours? We find that at this level of volatility of demand, technology shocks account for roughly 40% of the observed volatility in output.

Now, returning to Figure 11 and Table 1, we can examine the performance of the SVAR procedure with the QDSVAR and the LSVAR specifications at this setting of demand and technology shocks. We see that the impact error for the QDSVAR specification is −300% and that this specification confidently rejects the possibility that the impact coefficient is positive (Figure 11). At this level of volatility the impact error for the LSVAR is 118% (Table 1), but clearly the confidence bands for the LSVAR procedure are so wide that the procedure cannot distinguish among models of interest.

Some intuition for why technology shocks with unit roots do not generate much volatility in hours comes from examining the static first-order condition for labor supply, which in our model is given by

$$c_t v'(l_t) = (1 - \tau_t)\theta \frac{y_t}{l_t}.$$

When the technology shock has a unit root, consumption c_t rises by about the same amount as output y_t in response to a technology shock, so that the labor supply l_t does not change much. Thus, as we have seen, to generate significant volatility in hours in this model, the volatility of the demand shock must be sizeable.

We also conducted a similar exercise with the investment wedge model. When the standard deviation of investment tax shocks is set so that this model reproduces the observed volatility in hours, as Table 2 shows, technology shocks account for 36% of the variability in output, the error in the QDSVAR impact coefficient is −296%, and the error in the LSVAR impact coefficient is almost as large, 276%.

[9] We use data on per capita hours for the U.S. economy as a whole kindly provided by Edward Prescott and Alexander Ueberfeldt.

D. Evidence Based on Maximum Likelihood Estimation

A fourth type of evidence on the relative sizes of the two shocks is the results of maximum likelihood estimation. We ask what relative sizes this procedure produces for various specifications of the observed variables. We conducted several such exercises and found that the estimates are sensitive to the list of observed variables: their range is enormous, and so are the impact errors of both specifications. The maximum likelihood estimates do not support the notion that technology shocks dominate demand shocks.

In our maximum likelihood procedure, we fix all the parameters of the model except for those of the stochastic processes. We then use the maximum likelihood procedure described by McGrattan (1994) and Anderson et al. (1996) to estimate the parameters of the vector AR1 process, (14) and (15), using several specifications for the observed variables, denoted V_t. In this procedure, we write the system in a state space form with a state of the form $X_t = (\log \hat{k}_t, \log z_t, \tau_{lt}, \log \hat{k}_{t-1}, \log z_{t-1}, \tau_{lt-1})$. The transition equation is

$$(38) \quad X_{t+1} = EX_t + F\hat{\varepsilon}_{mt+1},$$

where $\hat{\varepsilon}_{mt} = (0, \varepsilon'_{mt}, 0)$.The observer equation is

$$(39) \quad Y_t = HX_t.$$

We report on two specifications of the observer equation for both the QDSVAR and the LSVAR. The estimates of the key parameters and some statistics of interest for the two specifications are reported in Table 1. In the *hours specification*, we let the observed variables be $Y_t = (\Delta \log y_t,\ \log l_t)'$. In the *investment specification,* we let $Y_t = (\Delta \log y_t,\ \Delta \log x_t)'$. In both specifications, we impose an upper bound of .995 on the persistence parameter ρ_l. In the hours specification, the variability of output due to technology is fairly large, 76%; the impact error for the QDSVAR is -86%; and the impact error for the LSVAR is 3%. In the investment specification, the variability of output due to technology is more modest, 30%; the impact error for the QDSVAR is -438%; and the impact error for the LSVAR is 190%. Clearly, the impact error for both the QDSVAR and the LSVAR depends sensitively on the specification of observed variables.

We then asked which specification is preferable, in the sense that it leads to more accurate estimates of the key parameters of the stochastic process. To answer this question, we conducted Monte Carlo experiments for our baseline business cycle model. We set the key parameters at $\rho_l = .99$, $\sigma_l = 1\%$, and $\sigma_z = 1\%$. We generated 1,000 simulations of the same length as the actual data. For each simulation, we estimated the parameters of the stochastic process with maximum likelihood using the specifications of the observed variables. We imposed the same bound on ρ_l of .995 as in our estimation using actual data.

>From Table 3 we see that the investment specification clearly yields more accurate estimates of the model parameters than does the hours specification. We repeated this exercise using higher values of ρ_l and found that the investment specification continues to yield more accurate estimates of the model parameters. These findings lead us to prefer the investment specification for estimating the model's parameters.

Clearly, the variability of output due to technology shocks associated with the maximum likelihood estimates is sensitive to the variables included in the observer equation, especially investment. The reason for this sensitivity is that a stripped-down model like ours cannot mimic well all of the comovements in U.S. data, so that it matters what features of the data the researcher is primarily interested in. Full information methods like maximum likelihood turn out to be sensitive to details such as which variables are included in the estimation. Our Monte Carlo experiments lead us to prefer the investment specification. And this specification leads to a large impact error for the LSVAR.

We also used maximum likelihood to estimate the hours and the investment specifications for the investment wedge model. Here again we imposed an upper bound of .995 on the autoregressive parameter ρ_x. Table 2 shows that under both specifications, technology shocks account for about three-quarters of the variability in output. The associated impact error is about -70% for the QDSVAR and nearly 60% for the LSVAR. Given the size of the impact error, maximum likelihood estimates do not support the view that demand shocks are trivial.

E. Evidence from the Growth Model Literature

Finally, we consider one more type of evidence on the relative size of the two shocks: the business cycle literature based on the growth model. This literature contains a wide range of estimates for the fraction of output variability due to technology shocks.

The studies differ in their data and in the details of the procedure they use to compute estimates, but all attempt to measure a broadly similar conceptual object. And they get very different results. For example:

- Prescott (1986) computes the ratio of the variance of HP-filtered output in a real business cycle model with only technology shocks to the variance of HP-filtered output in U.S. data. He finds this ratio to be 76%.
- Eichenbaum (1991) uses generalized method of moments procedures on an estimated business cycle model and as a measure of this fraction uses the ratio of the model's variance for HP-filtered output with only technology shocks to the variance of HP-filtered output in the data. Eichenbaum finds that for his measure of this fraction, a reasonable range is extremely wide, from 5% to 200%.[10]
- McGrattan (1994) uses maximum likelihood procedures on an estimated business cycle model and as a measure uses the fraction of total variance explained by innovations in technology. For her measure, McGrattan reports a point estimate of 41% with a standard error of 46%, which suggests a wide range of uncertainty for this measure.

The message we get from these and related studies in the business cycle literature is that a plausible case can be made that in the U.S. data, technology shocks account for essentially any value between zero and 100% of output variance. Put differently, when the U.S. data are viewed through the lens of the growth model, dismissing any estimate in this range is unreasonable.

In sum, the evidence based on Galí's result, other SVAR literature, the actual volatility of U.S. hours worked, other estimation methods, and the growth model literature makes clear that the U.S. data do not definitively say that technology shocks account for virtually all of the movements in output. Indeed, serious research cannot ignore the possibility that other shocks play an important role. If they do, then according to our test, the common SVAR procedure is not useful in developing business cycle theories.

6. SVARs with Long-Run Restrictions in Practice

Thus far we have shown that the confidence bands of SVARs with long-run restrictions are large when confronted with data from our model, at least when demand shocks are non-trivial. We have also shown that these confidence bands are particularly large for the LSVAR

[10] In a summary of the evidence on this fraction, Eichenbaum eloquently states, "What the data are actually telling us is that, while technology shocks almost certainly play some role in generating the business cycle, there is simply an enormous amount of uncertainty about just what percent of aggregate fluctuations they actually do account for. The answer could be 70% as Kydland and Prescott (1989) [1991] claim, but the data contain almost no evidence against either the view that the answer is really 5% or that the answer is really 200%" (Eichenbaum 1991, p. 608).

compared to the QDSVAR. Here we conduct a different test: We examine the performance of the two SVAR specifications when confronted with actual U.S. data.

We find that whatever the data set or subsample, the QDSVAR specification produces basically the same results. That is not true for the LSVAR, however. For that specification, small conceptual differences in the underlying data which lead to small differences in cyclical properties lead to large differences in the impulse responses. The impulse responses of that specification are also very different across subsamples. These findings using actual data are consistent with our findings using data generated from our model, and they buttress a result of the rest of our work here, that the LSVAR specification is of questionable value in developing business cycle theory.

These findings come from applying the SVAR procedure to three popular U.S. data sets used in the SVAR literature: those of Francis and Ramey (2005b); Christiano, Eichenbaum, and Vigfusson (2003); and Galí and Rabanal (2005). The three data sets cover somewhat different time periods but use conceptually similar measures of productivity and hours worked. In Figure 12A, we plot the measures of hours used in the three studies. The figure suggests that the cyclical fluctuations of the three series are virtually identical, but that the series show some differences in trend behavior in the first part of the sample. In Figure 12B, we plot the HP-filtered cyclical component of these three series and see that they are indeed virtually identical.

The QDSVAR performs similarly with all three data sets. With all of them, a positive technology shock leads to a fall in hours on impact. Thus, here we focus mainly 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 specification yields sharply differing results for the three data sets. With this specification, on impact a positive technology shock leads to a fall in hours in one, a rise in hours in another, and basically no change in the third. These large differences in results across similar data sets are likely to be connected to our finding about the wide range of LSVAR impulse responses across simulations from our model.

The fall in hours is predicted by the LSVAR when we use the data that Francis and Ramey (2005b) constructed to estimate an LSVAR for the period 1948:1–2002:4. Their measure of productivity is the U.S. Bureau of Labor Statistics (BLS) series "Index of Output per Hour, Business." Francis and Ramey construct a new measure of hours by adjusting the

BLS series "Index of Hours in Business" for government employment and for demographic changes. Figure 12C illustrates that with these data, 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%, a value significantly different from zero at the 5% level.

The LSVAR predicts a rise in hours when 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 ($LBOUT$), and their measure of hours is business hours divided by the civilian population over the age of 16 (*LBMN* and *P16*). Figure 12D shows that with these data, 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 3 onward.[11]

Finally, the LSVAR predicts no change in hours on impact from the technology shock when we follow Galí and Rabanal (2005) and use data for 1948:1–2002:4. Their measure of productivity is business labor productivity, constructed as the ratio of nonfarm business sector output to hours worked by all persons in the nonfarm business sector. For hours, Galí and Rabanal use the ratio of nonfarm hours to the civilian population over the age of 16. The source is the Haver USECON database, and their measures of output, hours, and population are *LXNFO, LXNFH,* and *LNN,* respectively. Figure 12E indicates that with these data, a positive technology shock leads to a persistent but statistically insignificant rise in hours. On impact, the rise is essentially zero, and that is not significantly different from zero at the 5% level.

These sharply contrasting results have led researchers in the SVAR literature to draw sharply contrasting inferences. Francis and Ramey (2005b) argue that their evidence shows that real business cycle models are dead. Christiano, Eichenbaum, and Vigfusson (2003) maintain that the models are alive and well. Galí and Rabanal (2005) assert that the existing results are inconclusive; they prefer the alternative DSVAR specification, which, they argue, also shows that real business cycle models are dead. Interestingly, these studies use

[11] 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 10D with Figure 2 in their paper reveals that the mean impulse response is similar, but that they have much tighter confidence bands than we do.

similar conceptual measures of productivity, and two of them (Christiano, Eichenbaum, and Vigfusson 2003 and Galí and Rabanal 2005) use similar conceptual measures of hours as well.

A recent literature has argued that the LSVAR impulse responses are unstable in the sense that they differ across subsamples (Fernald 2005, Gambetti 2005, and the references in both). We find some evidence of instability as well. For example, with the Francis and Ramey (2005b) data set, the impact coefficient over the whole sample is -1.80 with a confidence band of $(-2.31, -.48)$, whereas over the period 1970:1–2002:4, it is quite different: .19 with a confidence band of $(-.55, .57)$.

The sensitivity of the LSVAR results to seemingly minor differences in measuring productivity and hours and across subsamples raises serious doubts about the reliability of the LSVAR procedure for drawing inferences about underlying models.

7. Related Literature

Our critique of the SVAR approach adds to nearly 30 years of other critiques of this approach. Previous critiques can be broadly divided into those based on invertibility problems, those using economic models as tests, those of circular specification searches, and those 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 Fernández-Villaverde, Rubio-Ramírez, and Sargent 2005.) Hansen and Sargent 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 although such problems may arise for some examples, they typically have not arisen in most applied models. Blanchard and Quah also argue that even when such problems do arise, the resulting inference mistakes may not be quantitatively large. Our critique is different from the Hansen-Sargent invertibility critique because our specifications do not suffer from invertibility problems.

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 models' impulse responses. Another difference is that we focus on the central conclusion of the recent SVAR literature, that technology shocks lead to a fall in hours, whereas Cooley and Dwyer focus on a variety of other issues. (For work similar in spirit to that of Cooley and Dwyer, see also McGrattan 2005.)

Erceg, Guerrieri, and Gust (2004) also test the SVAR procedure using economic models. In contrast to our focus on 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 that "overall, Galí's methodology appears to offer a fruitful approach to uncovering the effects of technology shocks"(p. 4). We conclude the opposite.

Uhlig (2005) criticizes what he sees as the circularity of searching over specifications until a certain pattern is found and then arguing that the data show 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.

8. Conclusion

Simple data analysis techniques that reliably point toward quantitatively promising models can be highly useful in applied economic analysis. The SVAR literature seems to hold out hope that SVAR is such a technique. The common, long-run restriction branch of this literature has attracted a great deal of interest because it claims that the procedure can accurately distinguish between promising classes of models without having to take a stand on the details of other shocks, besides minimal features such as orthogonality.

Our study concludes that this claim is true in principle: if researchers had long enough data sets, then the SVAR procedure would accurately identify the model's impulse response with only minimal assumptions on the details of the other shocks. In practice, however, the claim is not true. When demand shocks play a substantial role, SVARs with long-run restrictions yield accurate estimates of the impulse responses only if the sum of the autoregressive coefficients in the VAR is close to that of those in the model. With the typical small number of lags in the VAR, these sums are not close in a model like ours. Since the length of available data sets requires that the VAR have a small number of lags, SVARs with long-run restrictions work poorly.

The SVAR claim is also true in principle if the number of observables is sufficiently greater than the number of shocks, so that the observables in the VAR can be inverted to uncover the state of the model. Our examples suggest, however, that this finding must be interpreted with caution.

We emphasize that our analysis is not a critique of SVARs in general. It is also not a critique of SVARs with long-run restrictions. It is only a critique of SVARs with long-run restrictions that use the common approach of comparing inappropriate objects, empirical and theoretical impulse responses. As Sims (1989) has argued and Cogley and Nason (1995) have shown, SVARs that instead compare logically comparable objects may be useful in developing business cycle theories.

Elsewhere (in Chari, Kehoe, and McGrattan, 2007), we have argued for the usefulness of another approach to developing business cycle theory: business cycle accounting. This approach has the same goal as the SVAR approach—to quickly shed light on which of a class of models is promising—but business cycle accounting suffers from fewer shortcomings.

References

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

Anderson, Evan W.; Hansen, Lars Peter; McGrattan, Ellen R.; and Sargent, Thomas J. 1996. 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, 171–252. 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, 1341–93. 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.

Chari, V. V.; Kehoe, Patrick J.; and McGrattan, Ellen R. 2005. Technical Appendix to "A Critique of Structural VARs Using Real Business Cycle Theory." Manuscript, Federal Reserve Bank of Minneapolis.

__________. 2007. Business Cycle Accounting. *Econometrica* 75 (3): 781–836.

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.

Eichenbaum, Martin. 1991. Real Business-Cycle Theory: Wisdom or Whimsy? *Journal of Economic Dynamics and Control* 15 (4): 607–26.

Erceg, Christopher J.; Guerrieri, Luca; and Gust, Christopher. 2004. Can Long-Run Restrictions Identify Technology Shocks? International Finance Discussion Papers 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.

Fernald, John. 2005. Trend Breaks, Long-Run Restrictions, and the Contractionary Effects of Technology Improvements. Working Paper 2005-21, Federal Reserve Bank of San Francisco.

Fernández-Villaverde, Jesús; Rubio-Ramírez, Juan F.; and Sargent, Thomas J. 2005. A, B, C's, (and D's) for Understanding VARs. Technical Working Paper 0308, National Bureau of Economic Research.

Fisher, Jonas D. M. 2006. The Dynamic Effects of Neutral and Investment-Specific Technology Shocks. *Journal of Political Economy* 114 (3): 413–51.

Francis, Neville, and Ramey, Valerie A. 2005a. Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited. *Journal of Monetary Economics* 52 (8): 1379–99.

__________. 2005b. Measures of Per-Capita Hours and Their Implications for the Technology-Hours Debate. Working Paper 11694, National Bureau of Economic Research.

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

Galí, Jordi; López-Salido, David J.; and Vallés, Javier. 2003. Technology Shocks and Monetary Policy: Assessing the Fed's Performance. *Journal of Monetary Economics* 50: 723–43.

Galí, Jordi, and Rabanal, Pau. 2005. Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data? *NBER Macroeconomics Annual 2004,* vol. 19, edited by Mark Gertler and Kenneth Rogoff, 225–88. Cambridge, Mass.: MIT Press.

Gambetti, Luca. 2005. Technology Shocks and the Response of Hours Worked: Time-Varying Dynamics Matter. Manuscript, Universitat Pompeu Fabra, Barcelona, Spain.

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, Part 2): S223–50.

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,* 77–119. Underground Classics in Economics. Boulder: Westview Press.

Hurwicz, Leonid. 1950. Least Squares Bias in Time Series. In *Statistical Inference in Dynamic Economic Models,* edited by Tjalling C. Koopmans, 365–83. New York: Wiley.

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

__________. 1991. Hours of Employment Variation in Business Cycle Theory. *Economic Theory* 1 (1): 61–81.

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. 2005. Overdifferencing VARs is OK. Manuscript, Universitat Pompeu Fabra.

McGrattan, Ellen R. 1994. The Macroeconomic Effects of Distortionary Taxation. *Journal of Monetary Economics* 33 (3): 573–601.

__________. 2005. Comment on Galí and Rabanal's "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data?" *NBER Macro-*

economics Annual 2004, vol. 19, edited by Mark Gertler and Kenneth Rogoff, 289–308. Cambridge, Mass.: MIT Press.

Prescott, Edward C. 1986. Theory Ahead of Business Cycle Measurement. *Federal Reserve Bank of Minneapolis Quarterly Review.* 10 (4, Fall), 9–22.

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 W. 1988. Sources of Business Cycle Fluctuations. *NBER Macroeconomics Annual 1988,* vol. 3, edited by Stanley Fischer, 111–48. Cambridge, Mass.: MIT Press.

Sims, Christopher A. 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 the American Statistical Association* 67 (337): 169–75.

__________. 1989. Models and Their Uses. *American Journal of Agricultural Economics* 71 (2): 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. 2005. What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. *Journal of Monetary Economics* 52 (2): 381–419.

Appendix A
Identification Using Long-Run Restrictions

Here we show that the two key assumptions of an SVAR along with the sign restriction identify the impulse responses of both variables in the VAR to a technology shock as summarized by the first columns of the matrices A_i.

Consider the system of four (independent) equations made up of (8) and (9) in the four unknowns of A_0. These can be written as

$$(40)\quad A_0(1,1)^2 + A_0(1,2)^2 = \omega_{11}$$

$$(41)\quad A_0(1,1)A_0(2,1) + A_0(1,2)A_0(2,2) = \omega_{12}$$

$$(42)\quad A_0(2,1)^2 + A_0(2,2)^2 = \omega_{22}$$

$$(43)\quad \bar{C}(1,1)A_0(1,2) + \bar{C}(1,2)A_0(2,2) = 0.$$

First suppose that $\bar{C}(1,1)$ is not equal to zero. Then we can manipulate equations (40)–(43) to obtain one equation in $A_0(2,2)$:

$$fA_0(2,2)^2 + \left[\omega_{11} - f^2A_0(2,2)^2\right]^{1/2}\left[\omega_{22} - A_0(2,2)^2\right]^{1/2} = \omega_{12},$$

where $f = -\bar{C}(1,2)/\bar{C}(1,1)$. This equation is quadratic in $A_0(2,2)^2$ and can be solved to obtain

$$(44)\quad A_0(2,2)^2 = \frac{\omega_{11}\omega_{22} - \omega_{12}^2}{\omega_{11} - 2f\omega_{12} + f^2\omega_{22}},$$

while the other elements of A_0 are given by

$$A_0(1,2) = -fA_0(2,2)$$

$$A_0(1,1)^2 = \omega_{11} - f^2A_0(2,2)^2$$

$$A_0(2,1)^2 = \omega_{22} - A_0(2,2)^2.$$

We then need to use (11), the sign restriction,

$$(45)\quad \bar{C}(1,1)A_0(1,1) + \bar{C}(1,2)A_0(2,1) > 0,$$

to pick the relevant roots of the quadratics. Using the definition of f, we can rewrite equation

(41) as

$$(46) \quad A_0(1,1)A_0(2,1) + fA_0(2,2)^2 = \omega_{12}$$

and use that to rewrite (45) as

$$(47) \quad \bar{C}(1,1)A_0(1,1) + \bar{C}(1,2)\left[\omega_{12} - fA_0(2,2)^2\right]/A_0(1,1) > 0.$$

Combined with (44), this equation pins down the sign of $A_0(1,1)$. Equation (46) pins down the sign of $A_0(2,1)$. Thus, we have shown that as long as $\bar{C}(1,1)$ is not equal to zero, the first column of A_0 is identified. Clearly, the first column of $A_i = C_i A_0$ is also identified for all $i \geq 1$.

When $\bar{C}(1,1)$ equals zero but $\bar{C}(1,2)$ does not, equation (43) implies that $A_0(2,2) = 0$, and a similar argument can be used to show that the first column of A_i is identified for $i \geq 0$.

Note that much of the literature does not explicitly mention that a necessary condition for the four conditions (40)–(43) to pin down the first column of A_0 is that at least one of $\bar{C}(1,1)$ or $\bar{C}(1,2)$ be nonzero. If both are zero, then (43) places no restrictions on A_0, and clearly the first column of A_0 varies in many solutions to the three equations (40)–(42) in the four unknowns of A_0. A condition that is sufficient to imply that at least one of $\bar{C}(1,1)$ or $\bar{C}(1,2)$ is nonzero is (10), so that a technology shock has a nonzero long-run effect on the level of labor productivity.

Appendix B
A Special Case of Proposition 2

Here we show that Fisher's (2006) model is not subject to our critique because it has a first-order autoregressive representation. We first lay out Fisher's (2006) model along the lines of Fernández-Villaverde, Rubio-Ramírez, and Sargent (2005) and then transform it to relate it to our baseline model. Finally, we work out the transformed model's state space representation and show that it falls under the domain of our Proposition 2. In this sense, it is not surprising that the common approach works well in Fisher's model.

In Fisher's model, the planner maximizes expected utility $E_0 \sum_{t=0}^{\infty} \beta^t [\log c_t^* + \psi \log(1 - l_t)]$ over per capita consumption c_t^* and per capita labor l_t, subject to the resource constraint

(48) $V_t c_t^* + x_t = A_t k_t^{\theta} {l_t}^{1-\theta}$

and the law of motion for capital

(49) $k_{t+1} = (1-\delta)k_t + V_t x_t,$

where x_t is investment, A_t is a neutral technology shock that follows the unit root process,

(50) $\log A_{t+1} = \mu_a + \log A_t + \log a_{t+1},$

and V_t is an investment-specific technology shock that follows the unit root process of the form

(51) $\log V_{t+1} = \mu_v + \log V_t + \log v_{t+1}.$

Note that V_t is also the price of the investment good relative to output goods, and output y_t is $A_t k_t^{\alpha} l_t^{1-a}$.

We can transform Fisher's model to be similar to that of our baseline model. Let $c_t = V_t c_t^*$ and substitute (49) into (48) to give

(52) $c_t + k_{t+1} - (1-\delta)k_t x_t = V_t A_t k_t^{\theta} {l_t}^{1-\theta} (\equiv \hat{y}_t).$

Letting $Z_t^{1-\alpha} = V_t A_t$ and noting that in the objective function $\log c_t^* = \log V_t + \log c_t$, this model is equivalent to the baseline model (with an irrelevant additive constant in the objective function). The log-linearized decision rules for Fisher's economy thus are (16)–(18) without the demand shock. Note that care must be taken not to confuse output in Fisher's model with output in the equivalent baseline model: in Fisher's model, again, output y_t is $A_t k_t^{\alpha} l_t^{1-a}$, so that $V_t y_t = V_t A_t k_t^{\theta} l_t{}^{1-\theta} = k_t^{\theta}(Z_t l_t)^{1-\theta} = Z_t \hat{y}_t$; therefore,

$$\text{(53)} \quad \Delta \log y_t = \Delta \log \hat{y}_t + \left(\frac{1}{1-\theta}\right) \log a_t + \left(\frac{\theta}{1-\theta}\right) \log v_t.$$

Fisher assumes that the observed variables are the growth rate of labor productivity $\Delta \log(y_t/l_t)$, the labor input l_t, and the change in the log of the relative price of investment $\Delta \log V_t$, which equals $\log v_t$. We can use the equivalent baseline model to work out the state space system for our model. Note from (53) that to recover $\log y_t$ we need to record both $\log a_t$ and $\log v_t$ separately in the state. The state of the system is $X_t = (\log \hat{k}_t, \log a_t, \log v_t)'$. The state equation $X_{t+1} = \hat{A} X_t + \hat{B} \varepsilon_{mt+1}$ is

$$\begin{bmatrix} \log \hat{k}_{t+1} \\ \log a_{t+1} \\ \log v_{t+1} \end{bmatrix} = \begin{bmatrix} \gamma_k & -\gamma_k/(1-\theta) & -\gamma_k/(1-\theta) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \log \hat{k}_t \\ \log a_t \\ \log v_t \end{bmatrix} + \begin{bmatrix} 0 \\ \log a_{t+1} \\ \log v_{t+1} \end{bmatrix}.$$

The observer equation $Y_t = \hat{C} X_t$ is

$$\begin{bmatrix} \Delta \log(y_t/l_t) \\ \log l_t \\ \log v_t \end{bmatrix} = \begin{bmatrix} \theta(1-a)(1-\frac{1}{\gamma_k}) & \frac{1-\theta(1-a)}{1-\theta} & \frac{\theta a}{1-\theta} \\ a & -\frac{a}{1-\theta} & -\frac{a}{1-\theta} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \log \hat{k}_t \\ \log a_t \\ \log v_t \end{bmatrix}.$$

Since $\hat{C}$ is invertible, Proposition 2 applies, and Fisher's model has a first-order autoregressive representation.

Table 1. Parameter Estimates and Statistics of Interest for the Model with Taxes on Labor

	Parameter Estimates			Statistics of Interest[a]		
					Impact Error	
Evidence	ρ_l	σ_z	σ_l	%var(y)	QDSVAR	LSVAR
Galí VAR response	.950	.0114	.0073	50	−220 (−344,−79)	76 (−230,245)
Hours volatility	.950	.0114	.0088	40	−300 (−448,−132)	118 (−252,322)
Maximum likelihood[b]						
Hours specification	.995 (.0093)	.0114 (.0006)	.0050 (.0005)	76	−86 (−171,−5)	3 (−219,123)
Investment specification	.942 (.0076)	.0178 (.0016)	.0173 (.0013)	30	−438 (−616,−226)	190 (−270,442)

[a] The first statistic is the variance of output due to the technology shock, reported as a percent. The last two are the mean impact errors for the QDSVAR and LSVAR specifications. The values in parentheses are means of the upper and lower means of 95 % confidence bands across 1,000 applications of the VAR procedure.

[b] For the maximum likelihood parameter estimates, the values in parentheses are standard errors. The hours specification uses observations on output and labor, and the investment specification uses observations on output and investment.

Table 2. Parameter Estimates and Statistics of Interest for the Model with Taxes on Investment

Evidence	Parameter Estimates			Statistics of Interest[a]		
					Impact Error	
	ρ_x	σ_z	σ_x	%var(y)	QDSVAR	LSVAR
Galí VAR response	.950	.0114	.0143	46	−221	188
					(−368,−67)	(−120,327)
Hours volatility	.950	.0114	.0175	36	−296	276
					(−471,−108)	(−88,431)
Maximum likelihood[b]						
Hours specification	.995	.0116	.0096	76	−73	57
	(.0071)	(.0006)	(.0010)		(−158,7)	(−145,137)
Investment specification	.995	.0088	.0071	77	−69	53
	(.0078)	(.0004)	(.0007)		(−152,8)	(−144,130)

[a] The first statistic is the variance of output due to the technology shock, reported as a percent. The last two are the mean impact errors for the QDSVAR and LSVAR specifications. The values in parentheses are means of the upper and lower means of 95 % confidence bands across 1,000 applications of the VAR procedure.

[b] For the maximum likelihood parameter estimates, the values in parentheses are standard errors. The hours specification uses observations on output and labor, and the investment specification uses observations on output and investment.

Table 3. Monte Carlo Analysis of Maximum Likelihood Estimation for Two Sets of Observables in the Model with Taxes on Labor

Estimates	Hours Specification[a]			Investment Specification[b]		
	ρ_l	σ_z	σ_l	ρ_l	σ_z	σ_l
True estimates	.990	.0100	.0100	.990	.0100	.0100
Monte Carlo estimates						
Mean	.980	.0101	.0096	.990	.0100	.0100
Maximum	.995	.0121	.0119	.992	.0117	.0128
Minimum	.838	.0083	.0074	.986	.0084	.0070
% Standard deviation	1.83	.053	.084	.076	.053	.083

[a] The hours specification uses observations on output and labor.

[b] The investment specification uses observations on output and investment.

Figure 1

Model Impulse Response of Hours to a Technology Shock

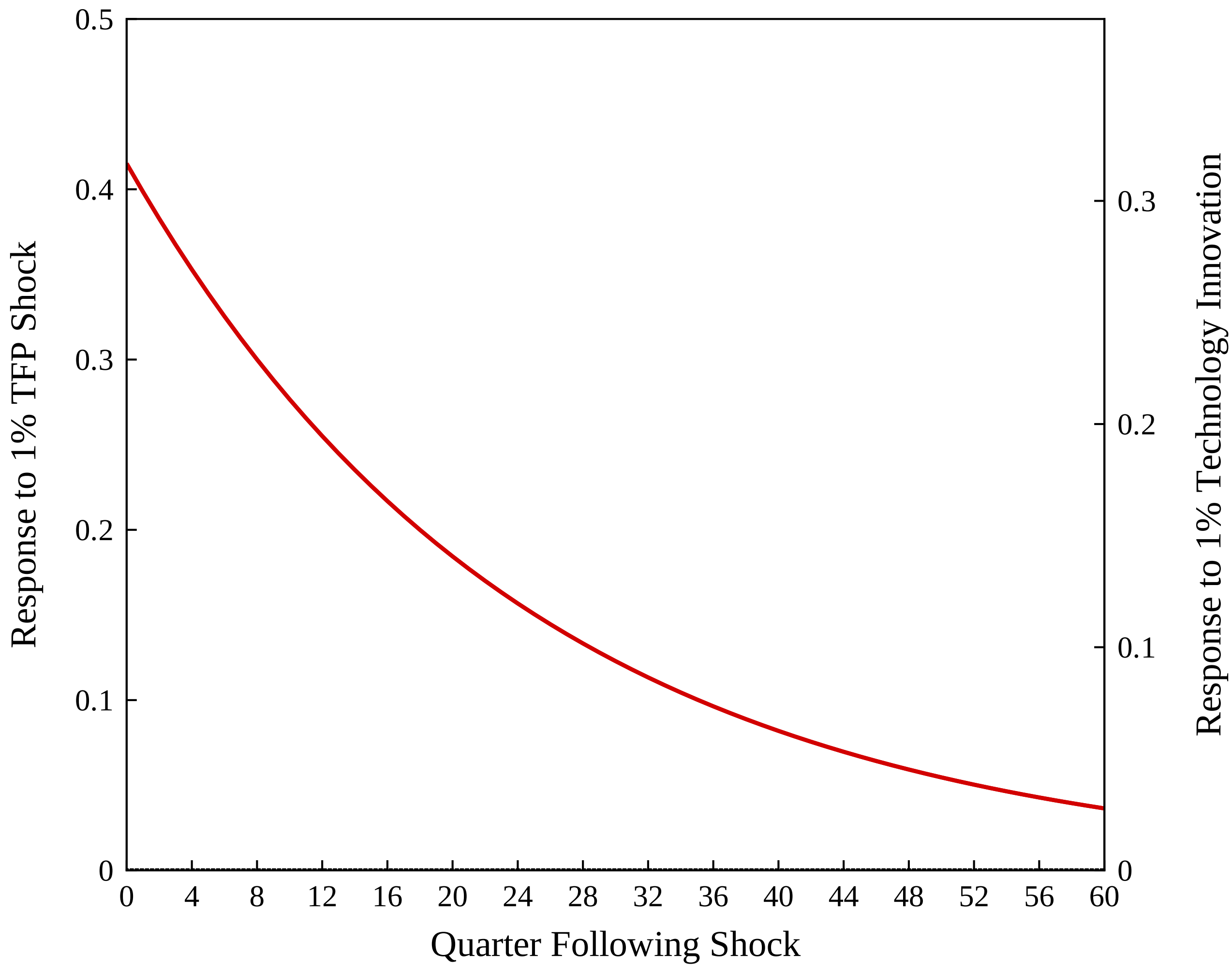


FIGURE 2

IMPACT ERRORS OF THE SVAR PROCEDURES

Mean Error in the Impact Coefficient of Hours From 1,000 Applications of the Four-Lag SVAR Procedures Applied to Model Simulations of Length 180, Varying Innovations of the Shock Processes

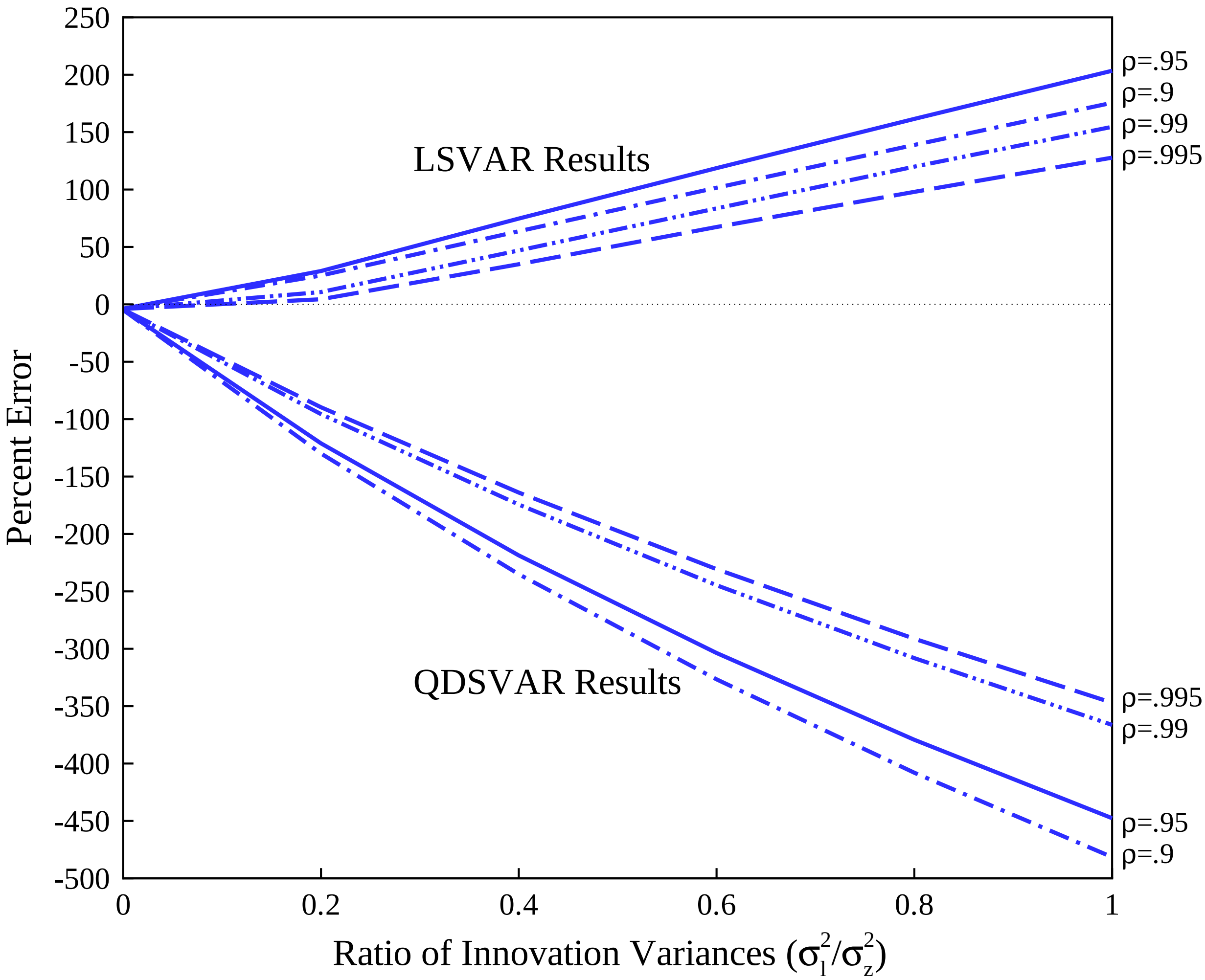


Figure 3a

Impact Errors and Bands of the QDSVAR Procedure

Mean Error in the Impact Coefficient of Hours (solid line) and 95% Confidence Bands (dashed lines) From 1,000 Applications of the Four-Lag QDSVAR Procedure with $\rho_l = .95$ Applied to Model Simulations of Length 180, Varying the Variance of Output Due to Technology

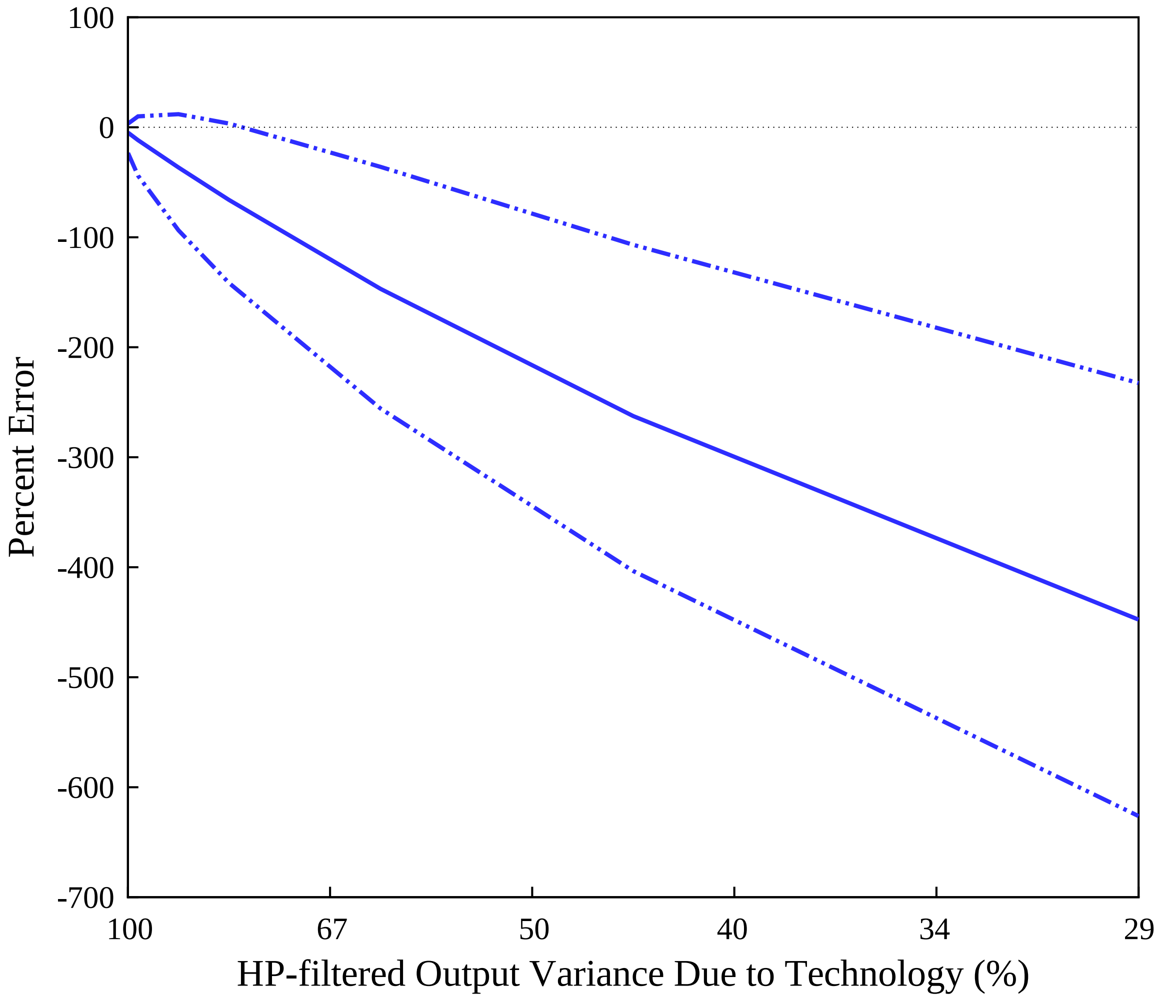


Figure 3b

Impact Errors and Bands of the LSVAR Procedure

Mean Error in the Impact Coefficient of Hours (solid line) and 95% Confidence Bands (dashed lines) From 1,000 Applications of the Four-Lag LSVAR Procedure with $\rho_l = .95$ Applied to Model Simulations of Length 180, Varying the Variance of Output Due to Technology

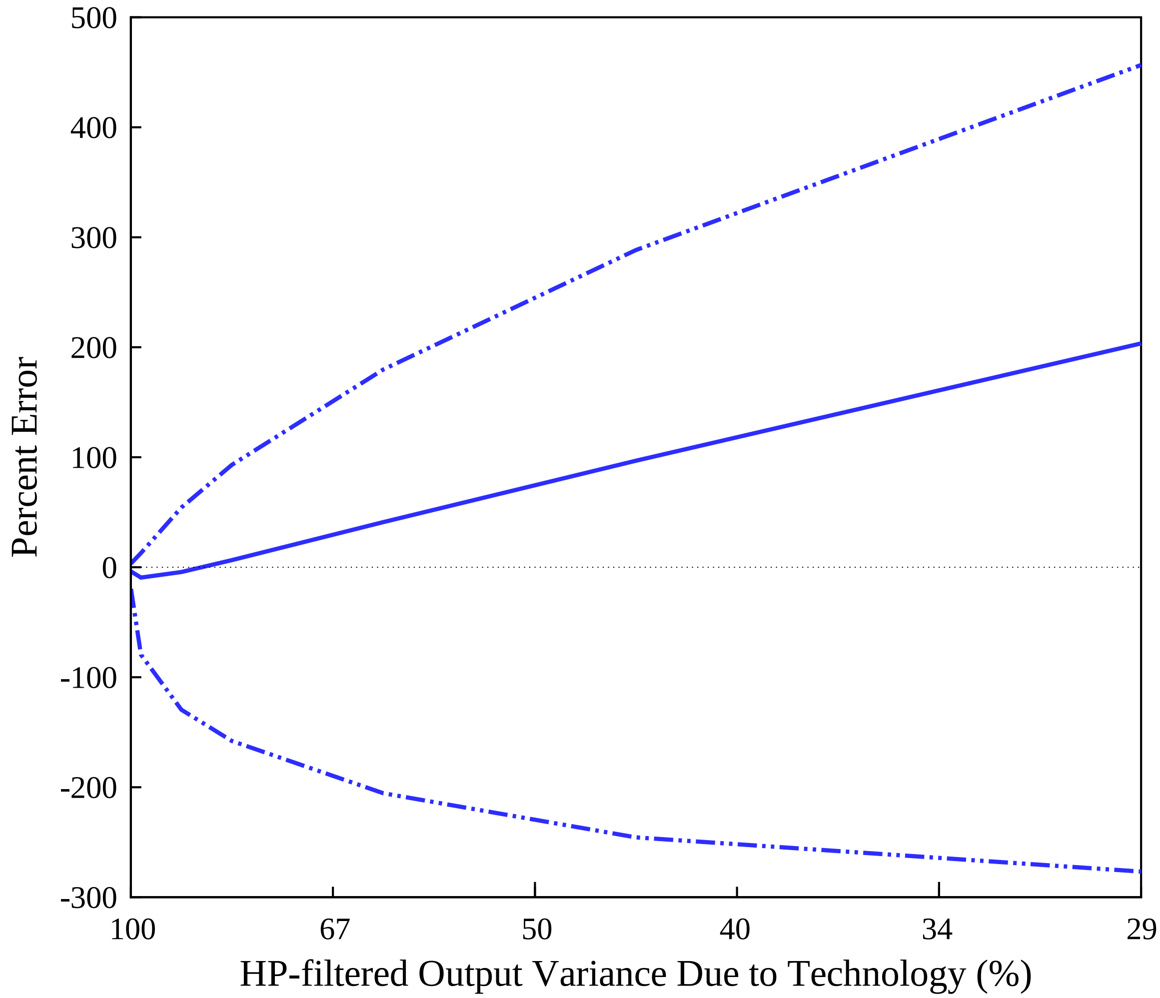
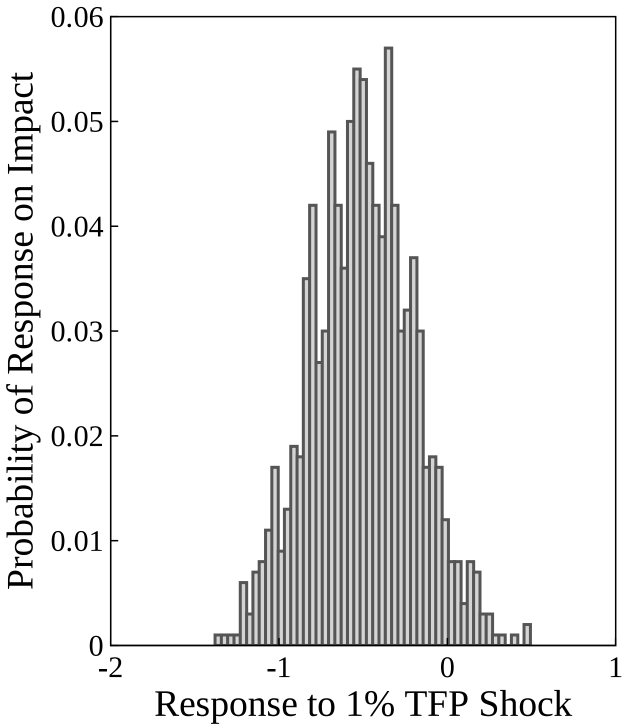


Figure 4a

QDSVAR Histogram and Bounds for Galí Parameters

Histogram of Impact Coefficient of Hours and 95% Bounds on Impulse Responses From 1,000 Applications of the Four-Lag QDSVAR Procedure to Model Simulations of Length 180



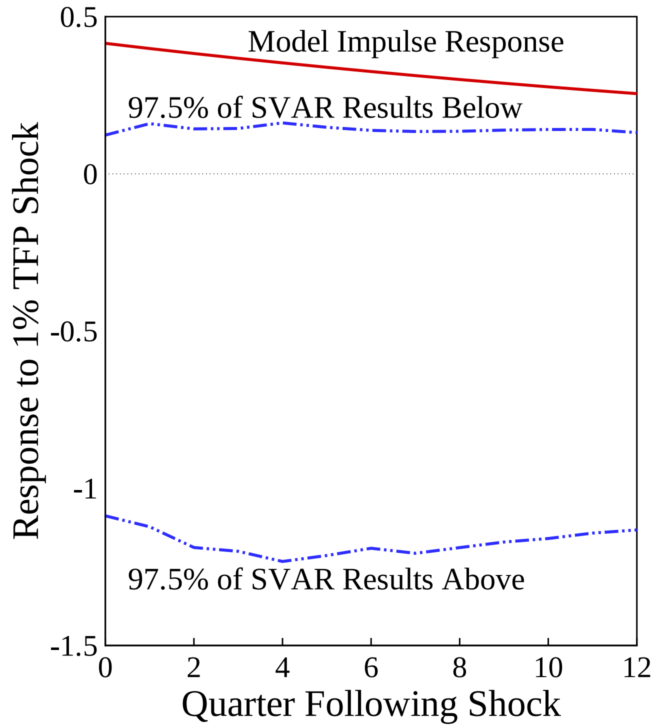


FIGURE 4B

QDSVAR RESPONSES AND BANDS FOR GALÍ PARAMETERS

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

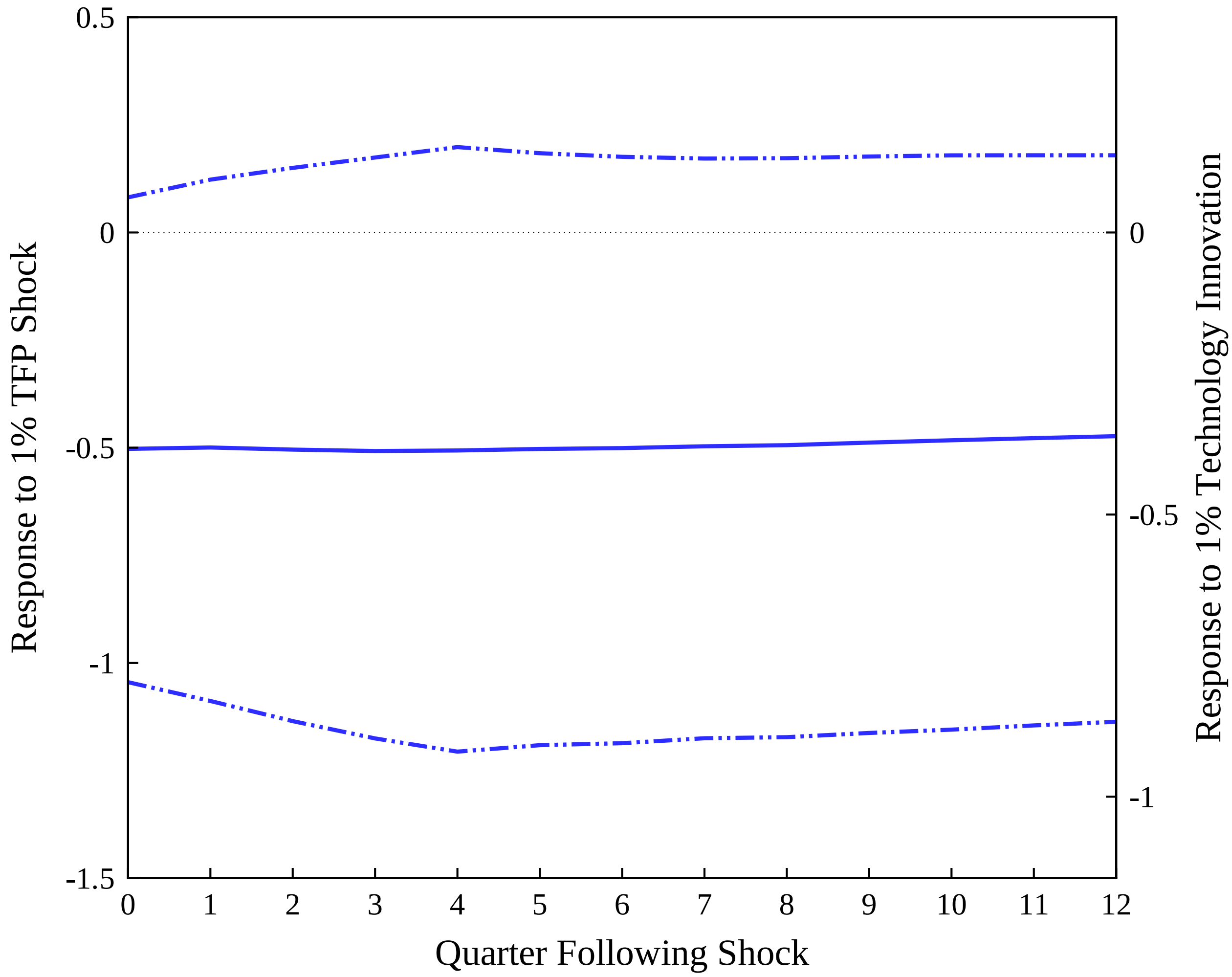


Figure 5a

LSVAR Histogram and Bounds for Galí Parameters

Histogram of Impact Coefficient of Hours and 95% Bounds on Impulse Responses From 1,000 Applications of the Four-Lag QDSVAR Procedure to Model Simulations of Length 180

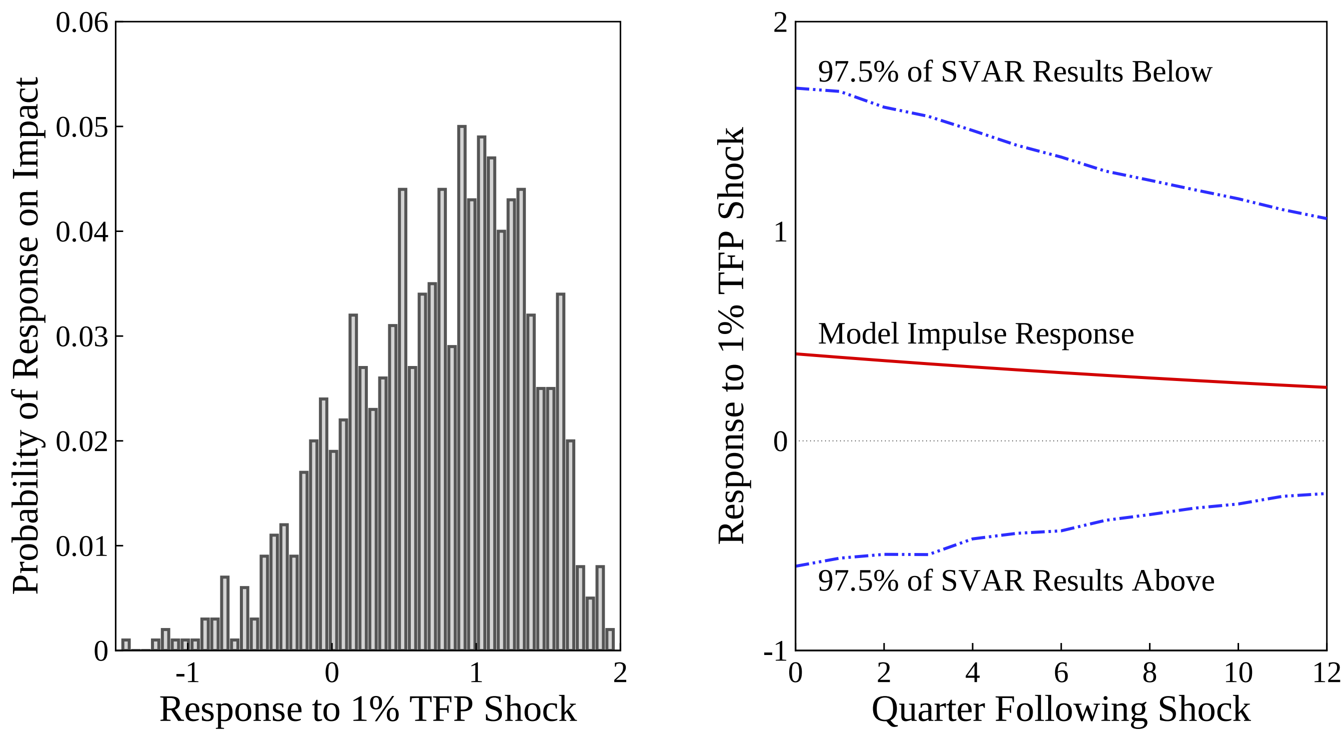


Figure 5b

LSVAR Responses and Bands for Galí Parameters

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

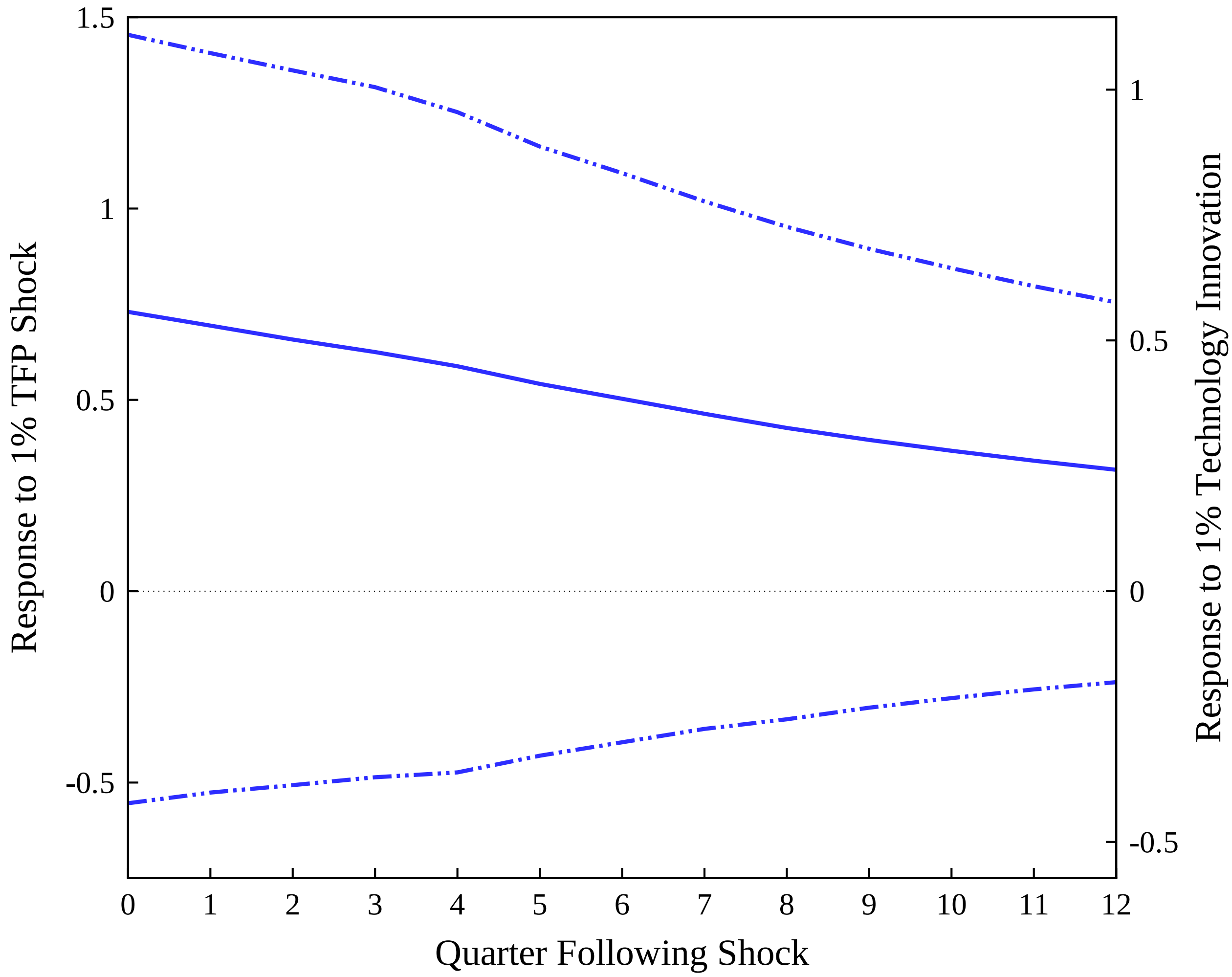


Figure 6a

Impact Errors and Bands of the QDSVAR Procedure

Mean Error in the Impact Coefficient of Hours (solid line) and 95% Confidence Bands (dashed lines) From 1,000 Applications of the Four-Lag QDSVAR Procedure with $\rho_l = .95$ Applied to Model Simulations of Length 180 and Population Errors (dotted line),

Varying the Variance of Output Due to Technology

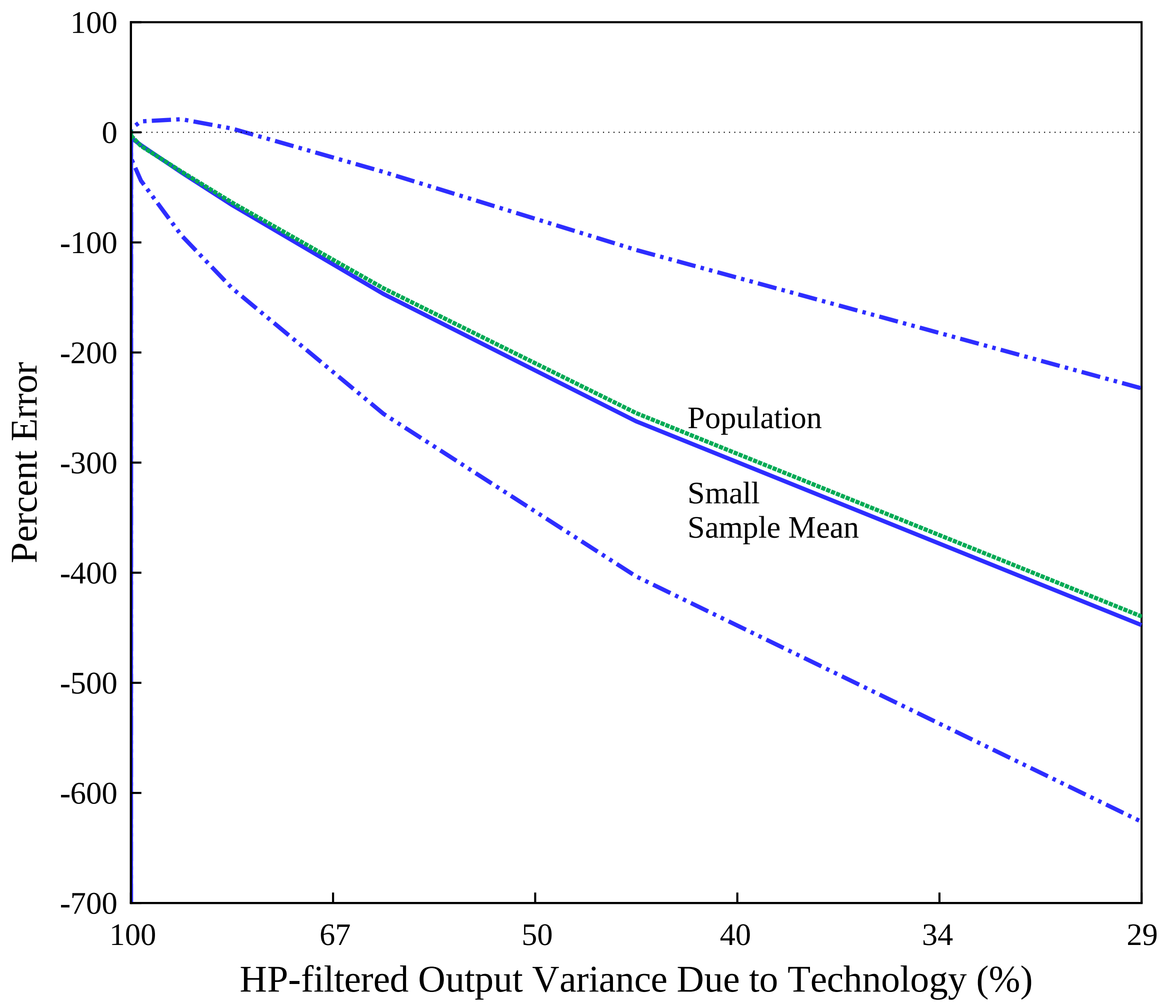


Figure 6b

Impact Errors and Bands of the LSVAR Procedure

Mean Error in the Impact Coefficient of Hours (solid line) and 95%
Confidence Bands (dashed lines) From 1,000 Applications of the Four-Lag
LSVAR Procedure with $\rho_l = .95$ Applied to Model Simulations
of Length 180 and Population Errors (dotted line),

Varying the Variance of Output Due to Technology

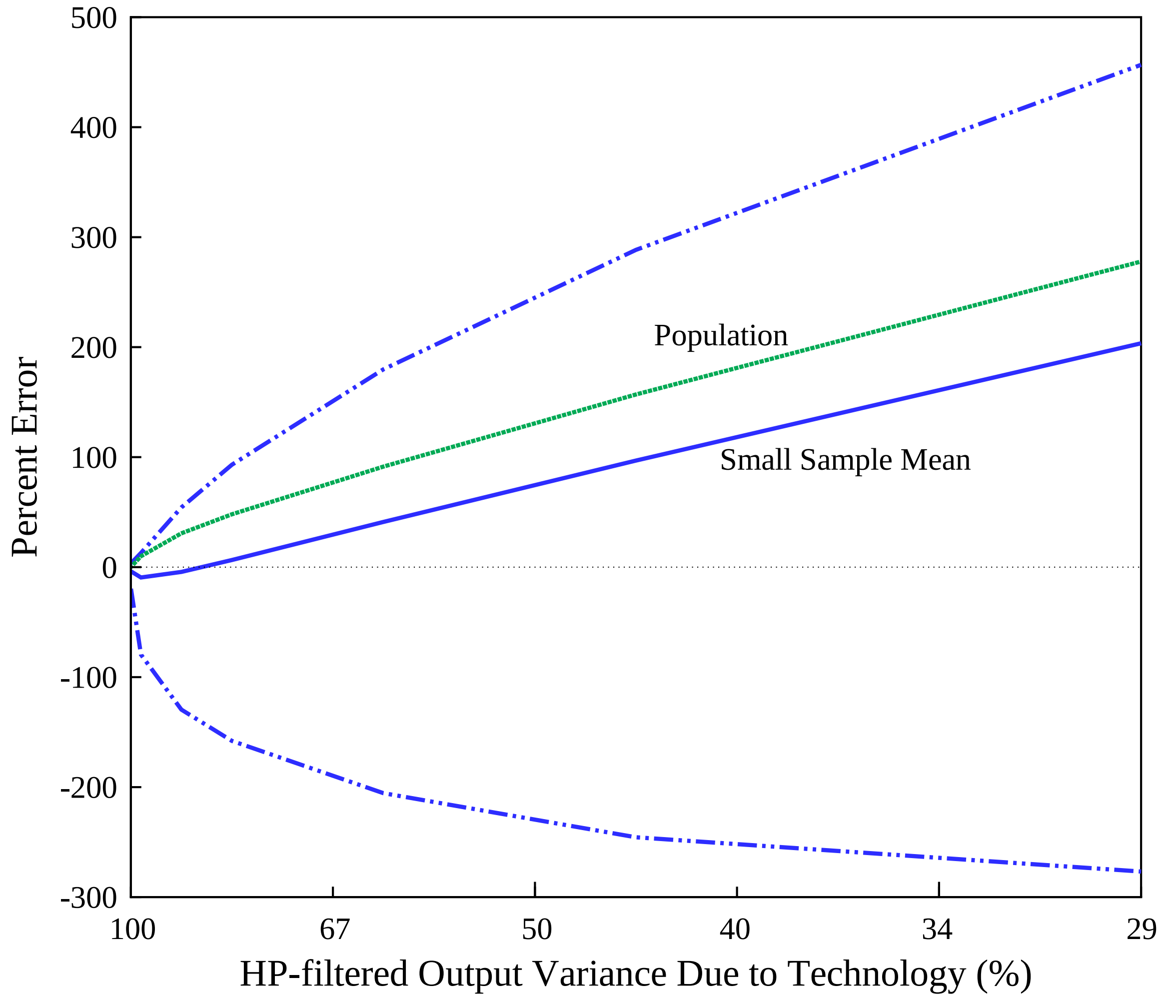
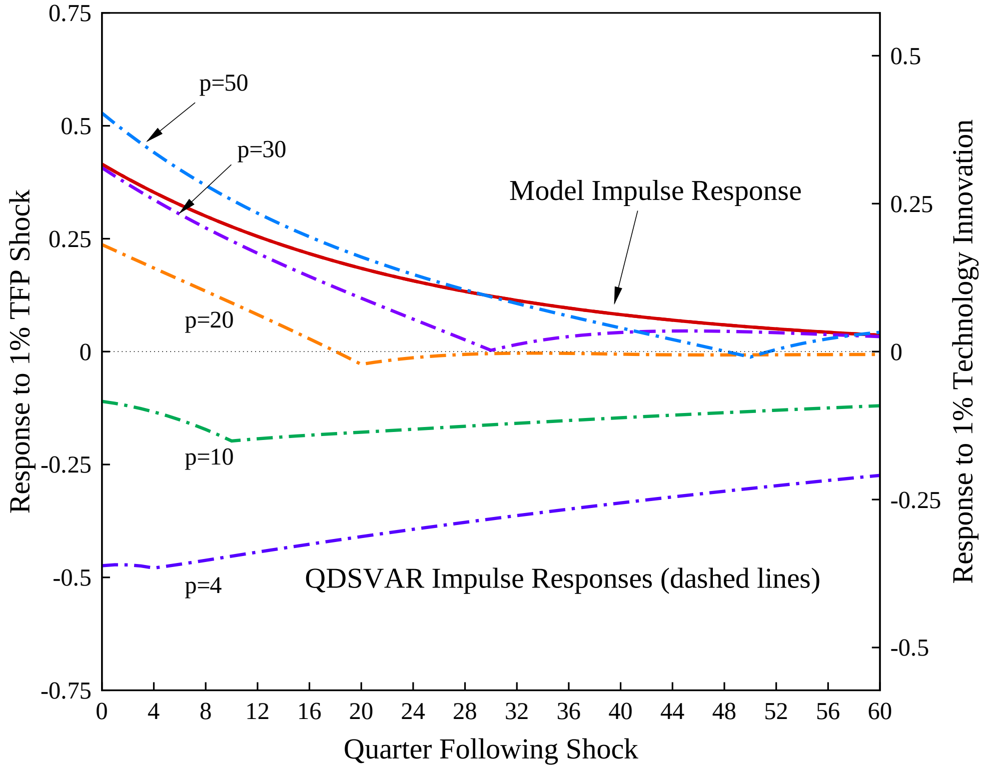


Figure 7

Model and QDSVAR Population Responses of Hours

Using Galí Parameters and Varying Lag Length p in QDSVAR Procedure



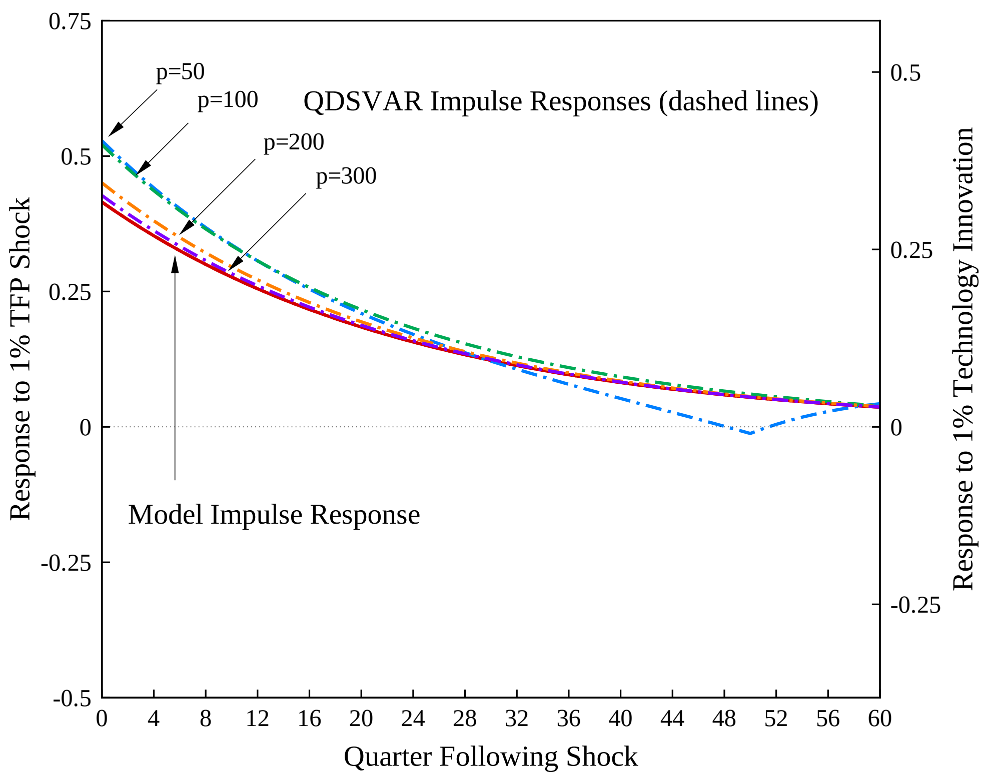


Figure 8

Model and LSVAR Population Responses of Hours

Using Galí Parameters and Varying Lag Length p in LSVAR Procedure

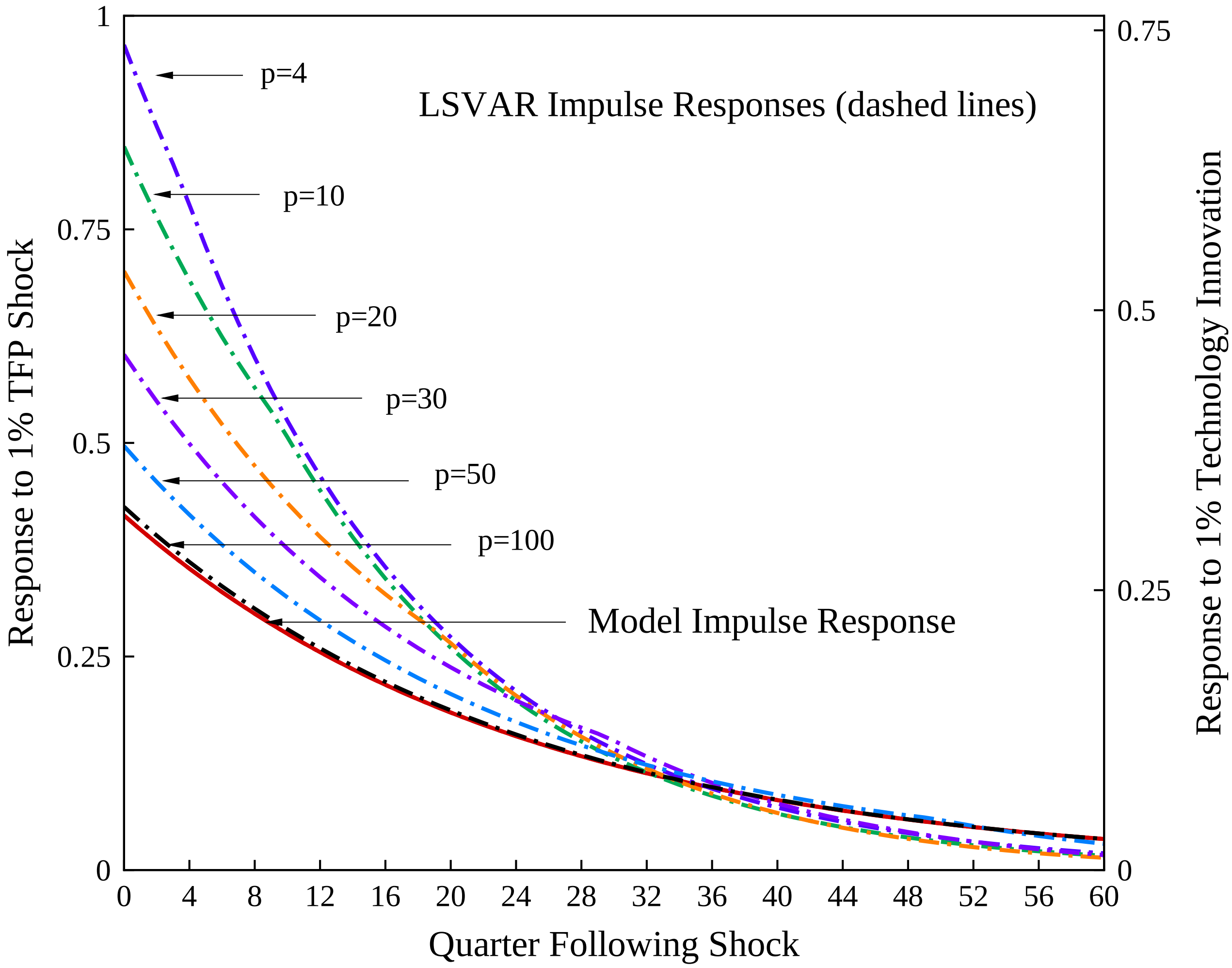


Figure 9a

Lag-Length Tests with Per Capita Hours Quasi-Differenced

Information Criteria for Tests of 1,000 Model Simulations of Length 180

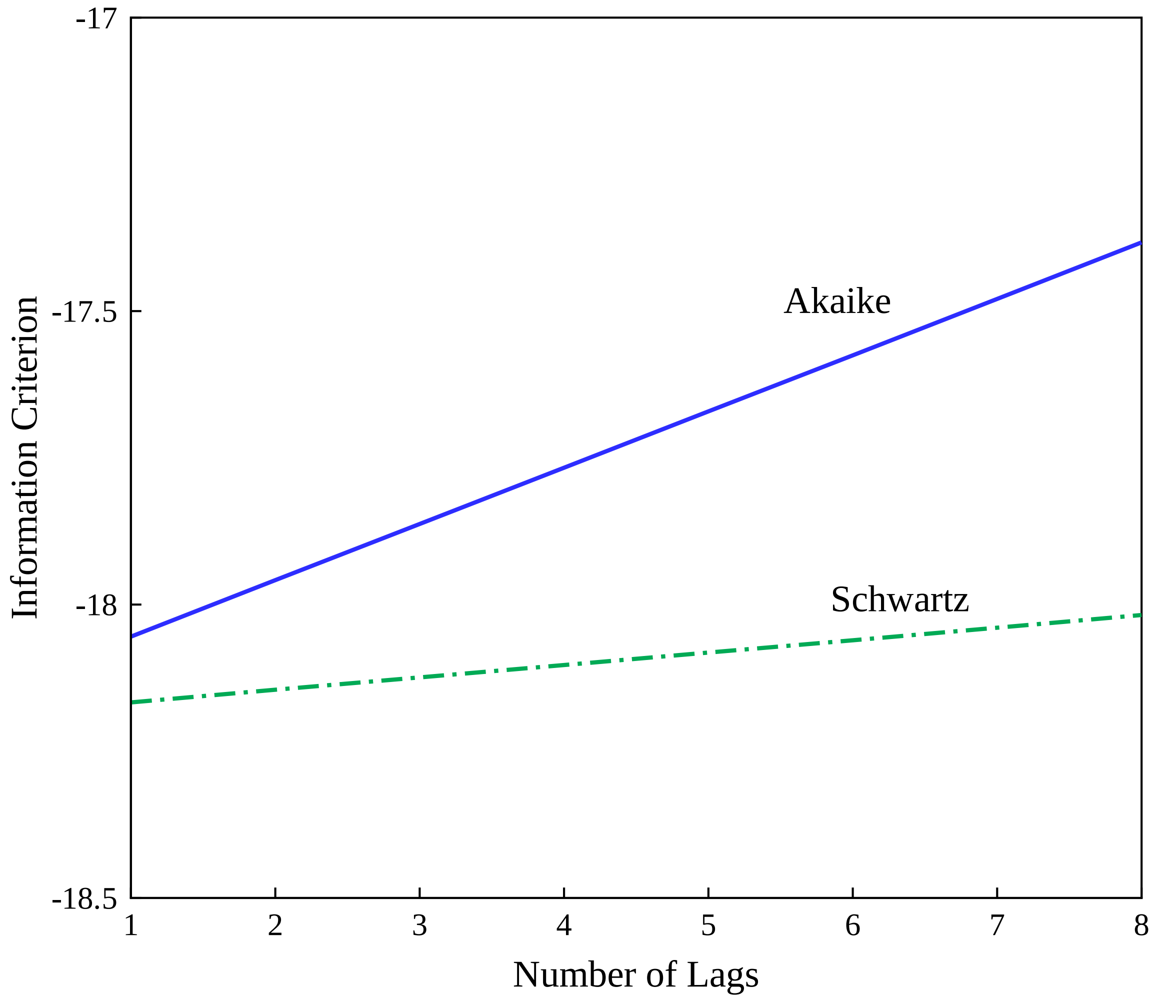


Figure 9b

Lag-Length Tests with Per Capita Hours in Levels

Information Criteria for Tests of 1,000 Model Simulations of Length 180

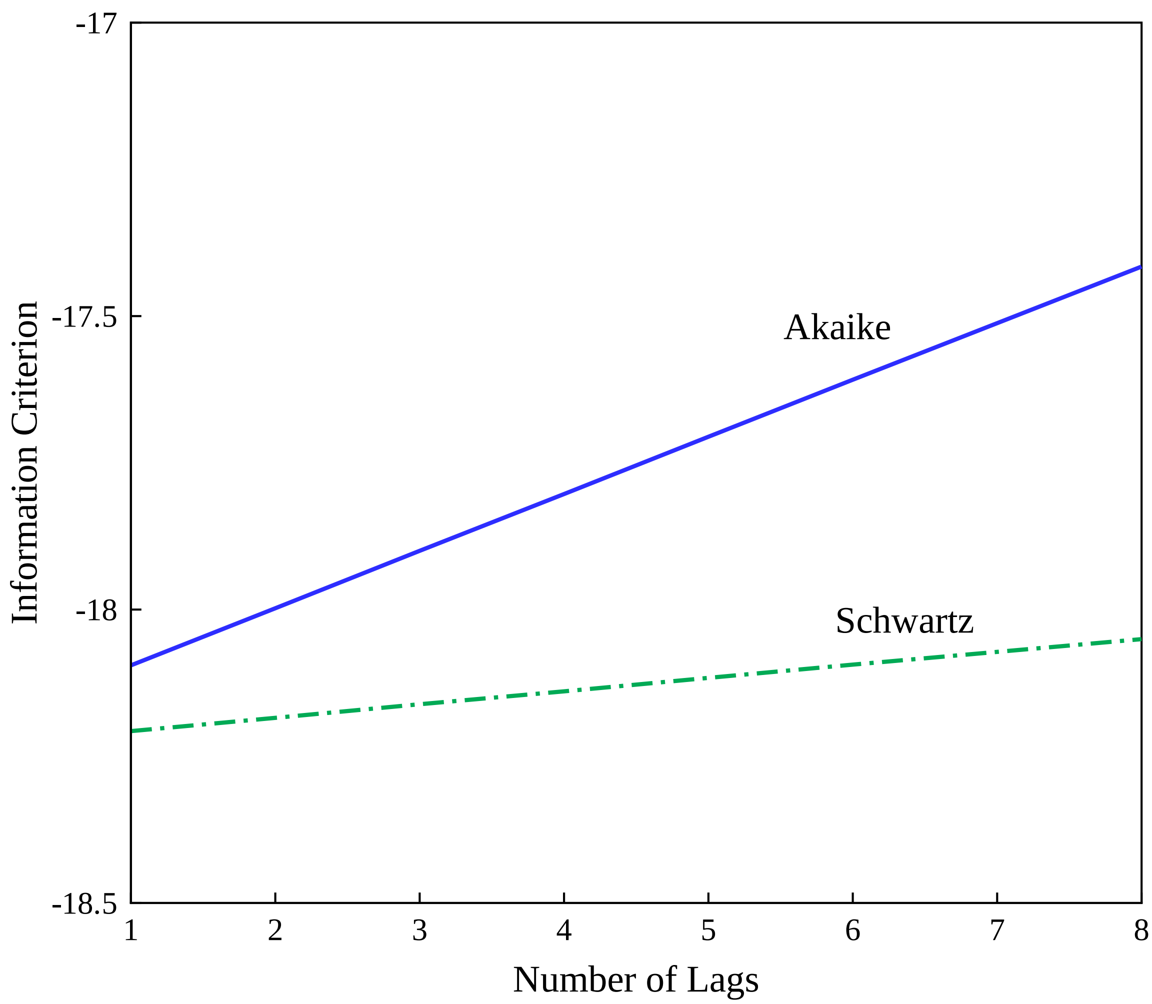


Figure 10

Impact Errors of the Three-Variable LSVAR Procedure

Mean Error in the Impact Coefficient of Hours (solid line)
and 95% Confidence Bands (dashed lines) From 1,000 Applications of
the Four-Lag, Three-Variable LSVAR Procedure Applied to Model
Simulations of Length 180 and Population Errors (dotted line),

Varying Innovations of the Shock Processes

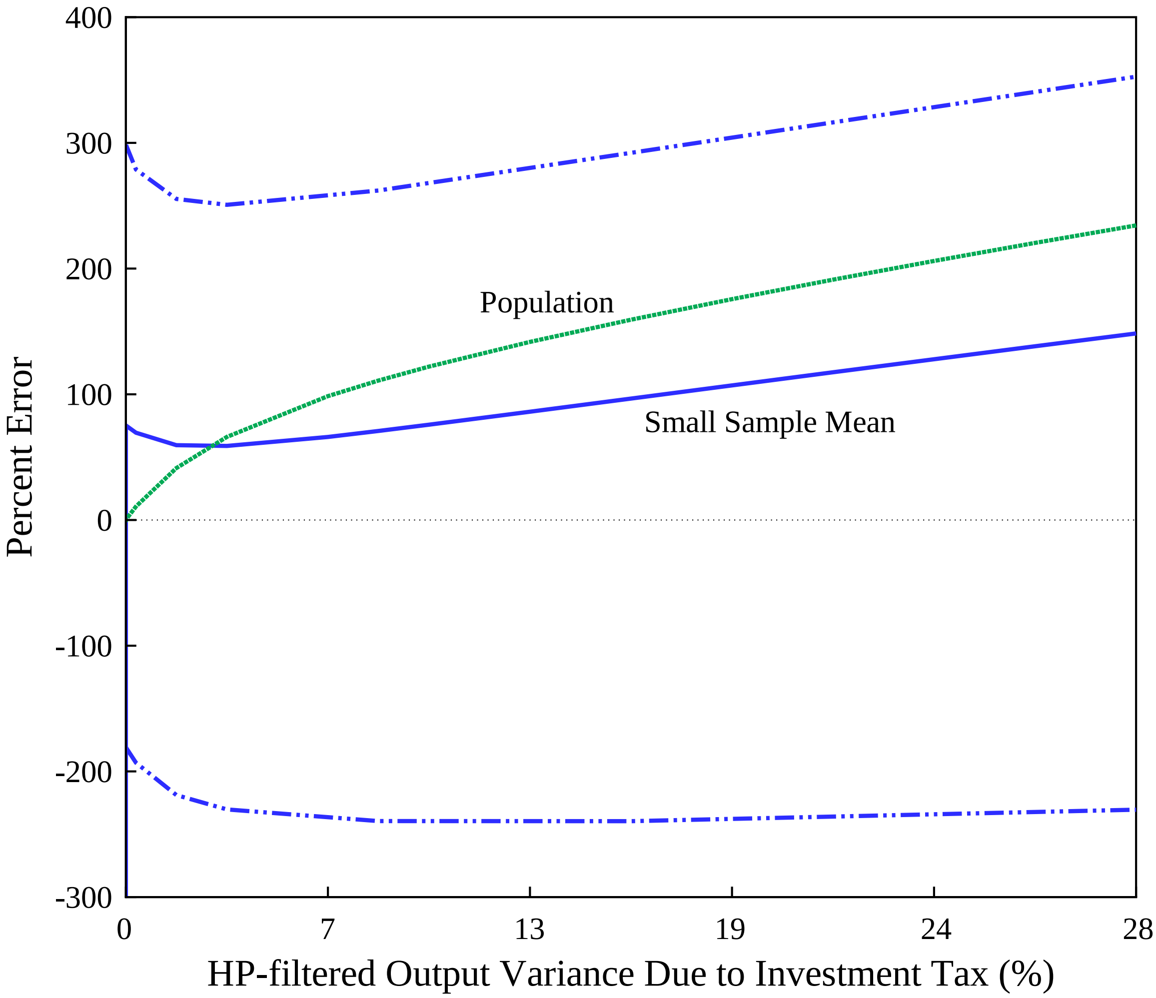


Figure 11

Innovation Variance Ratio Implied by Galí (1999) and U.S. Hours

Demand Shock Innovation is Varied to Reproduce Galí's (1999) Estimate for the Impulse Response of Hours to a Technology Shock and to Generate the Same Variance of Hours as in U.S. Data

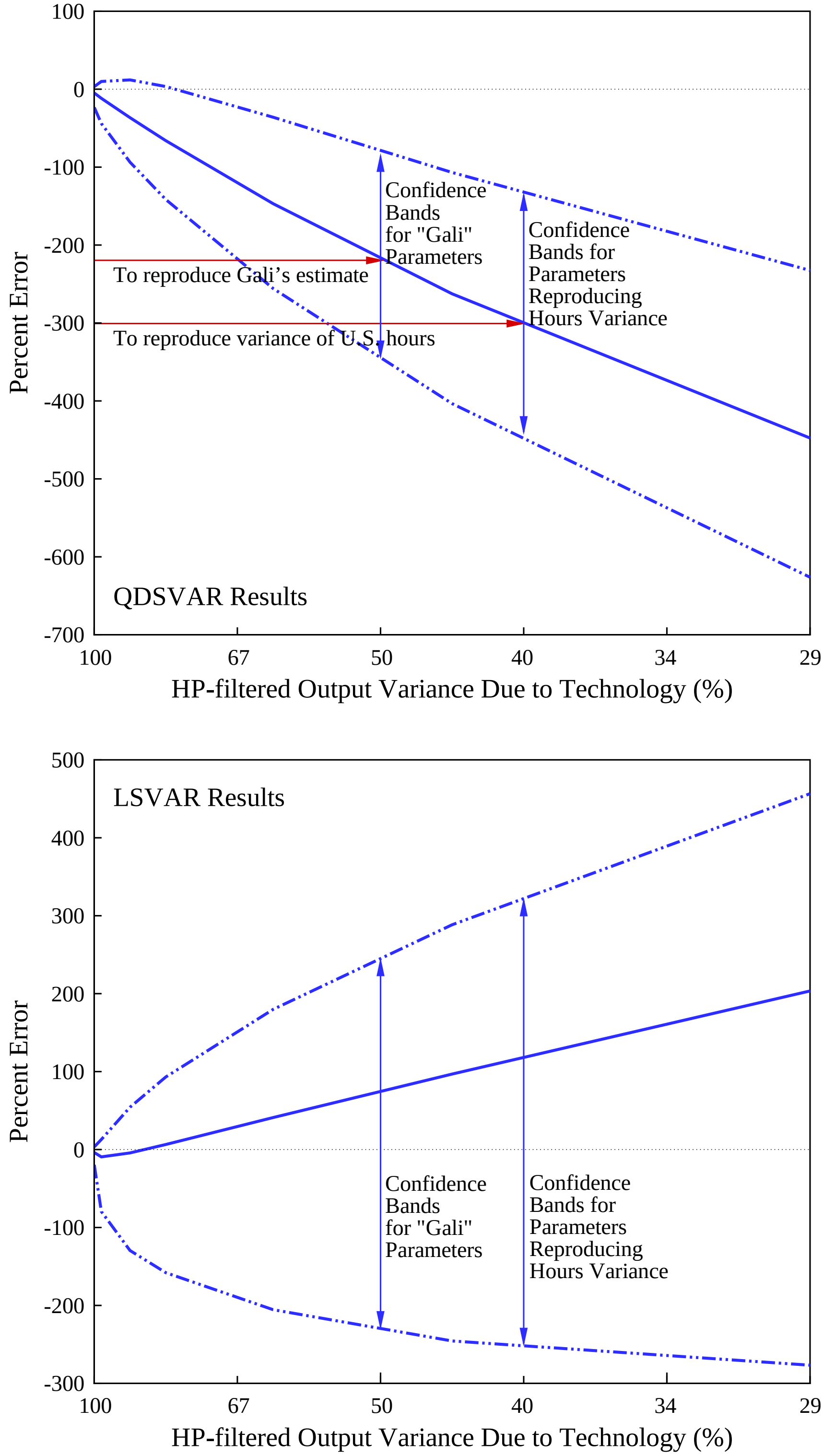


Figure 12a

Three Series for the Hours Per Capita Index

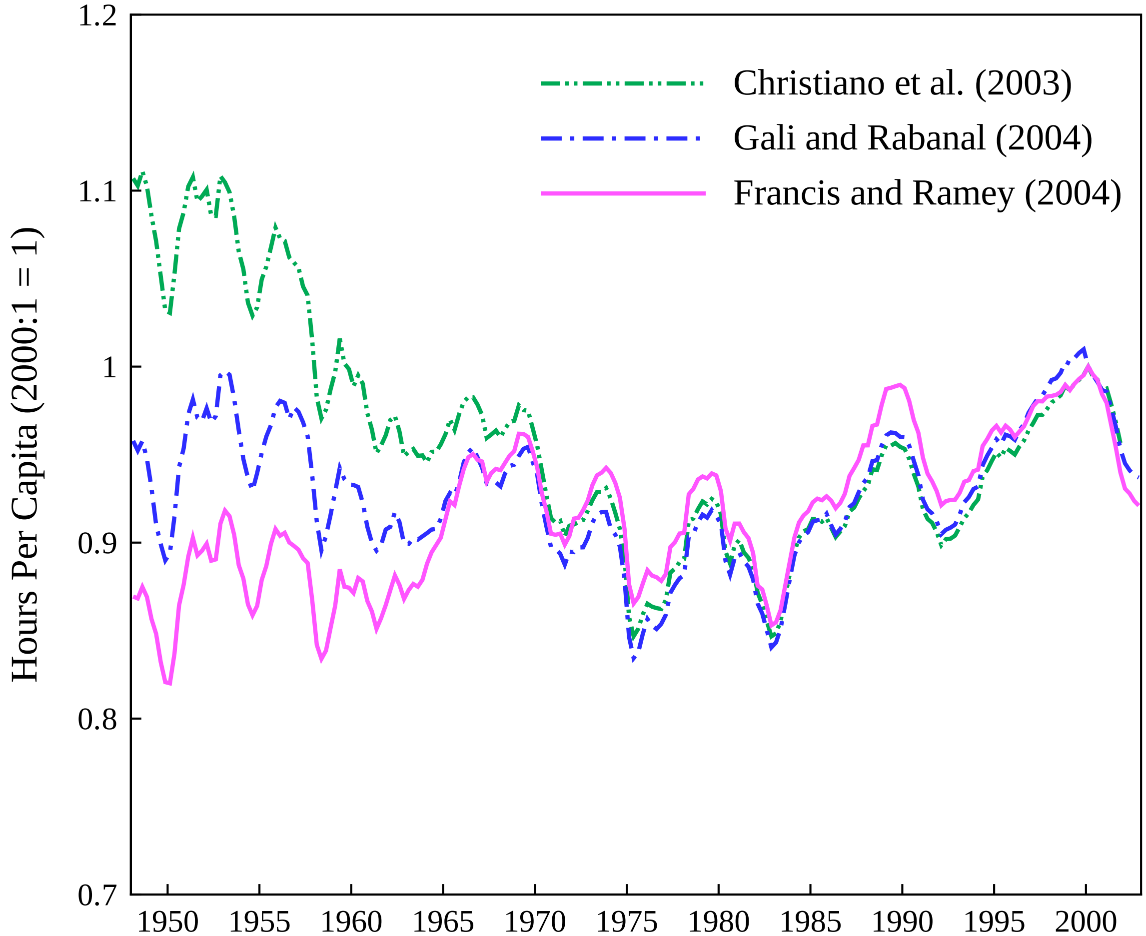


Figure 12b

Three HP-filtered Series for the Hours Per Capita Index

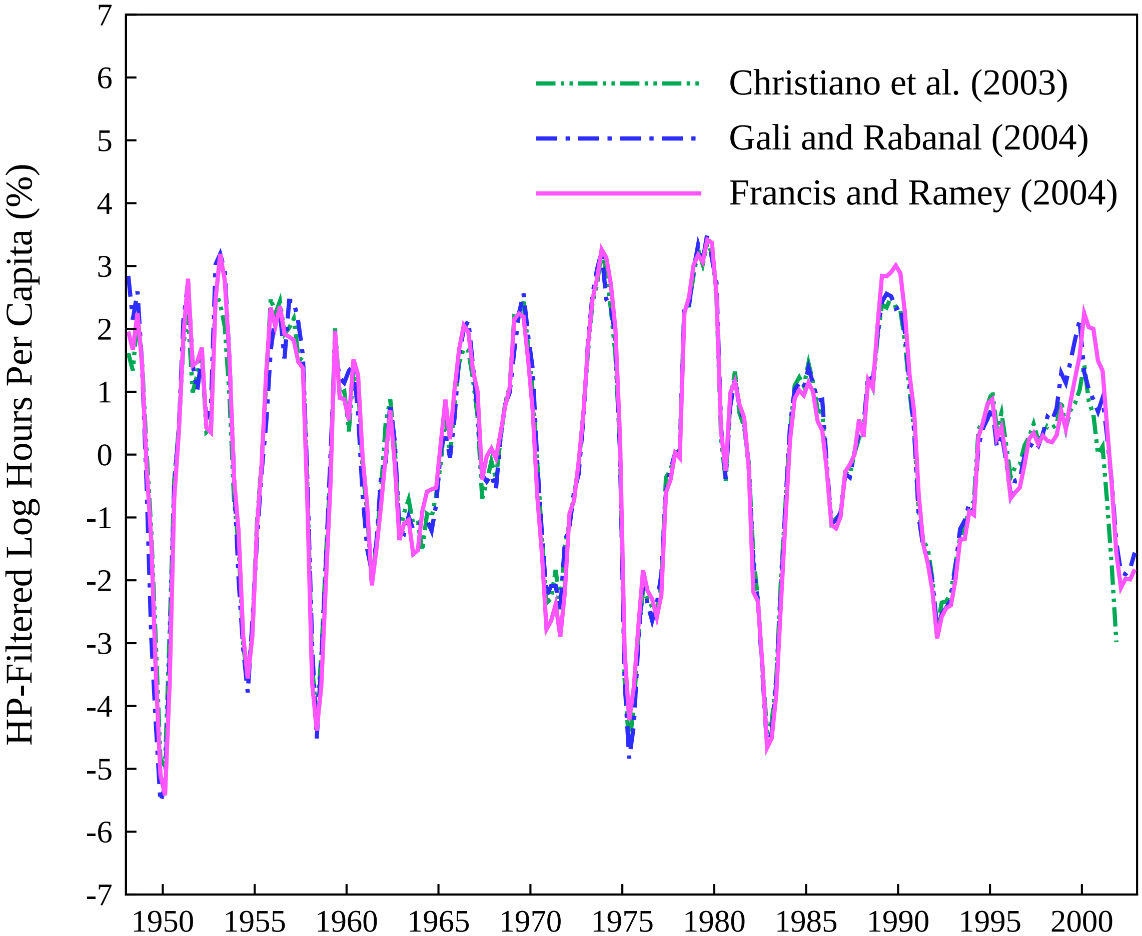


Figure 12c

Impulse Response for Francis and Ramey (2004) Data

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)

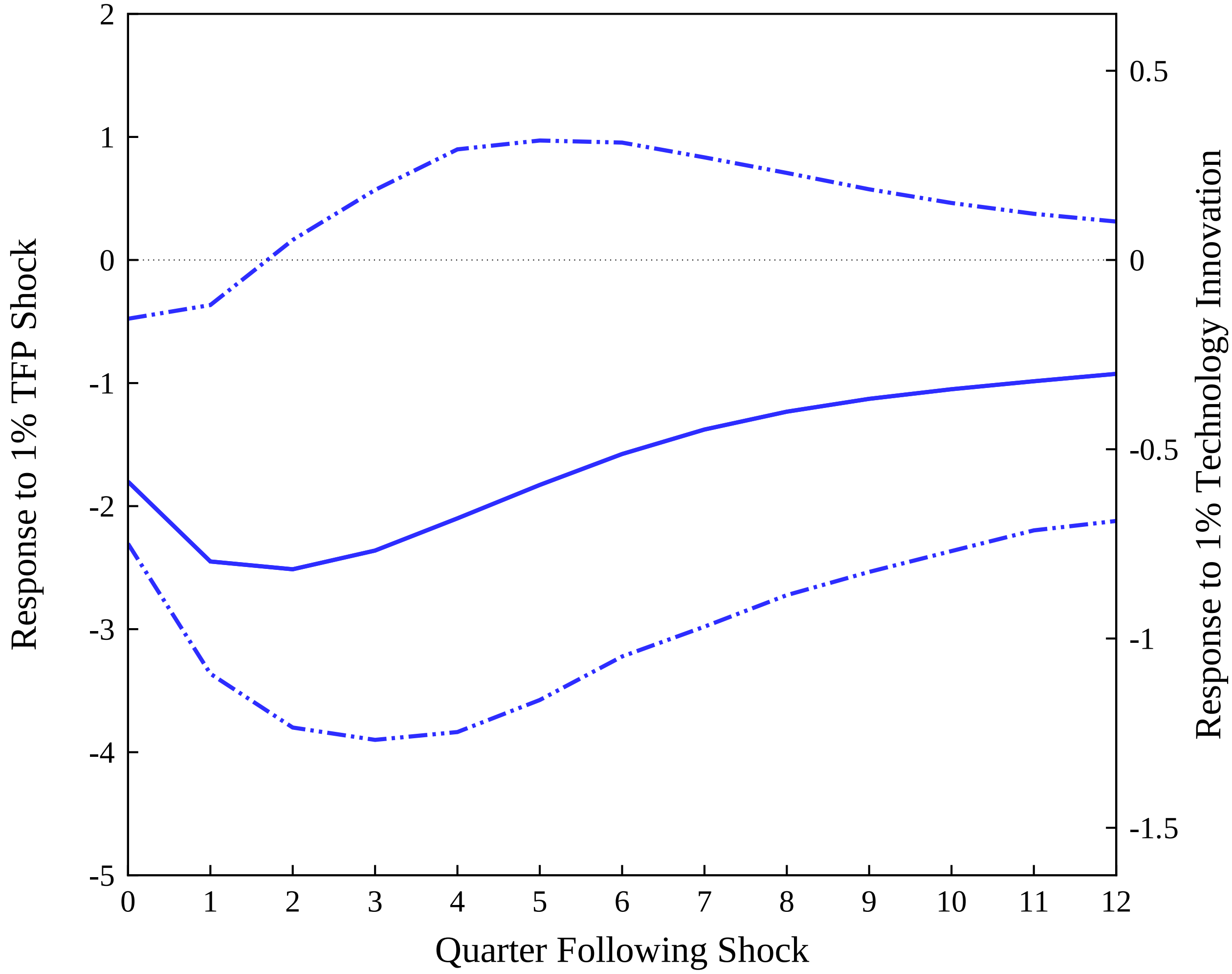


Figure 12d

Impulse Response for Christiano et al. (2003) Data

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)

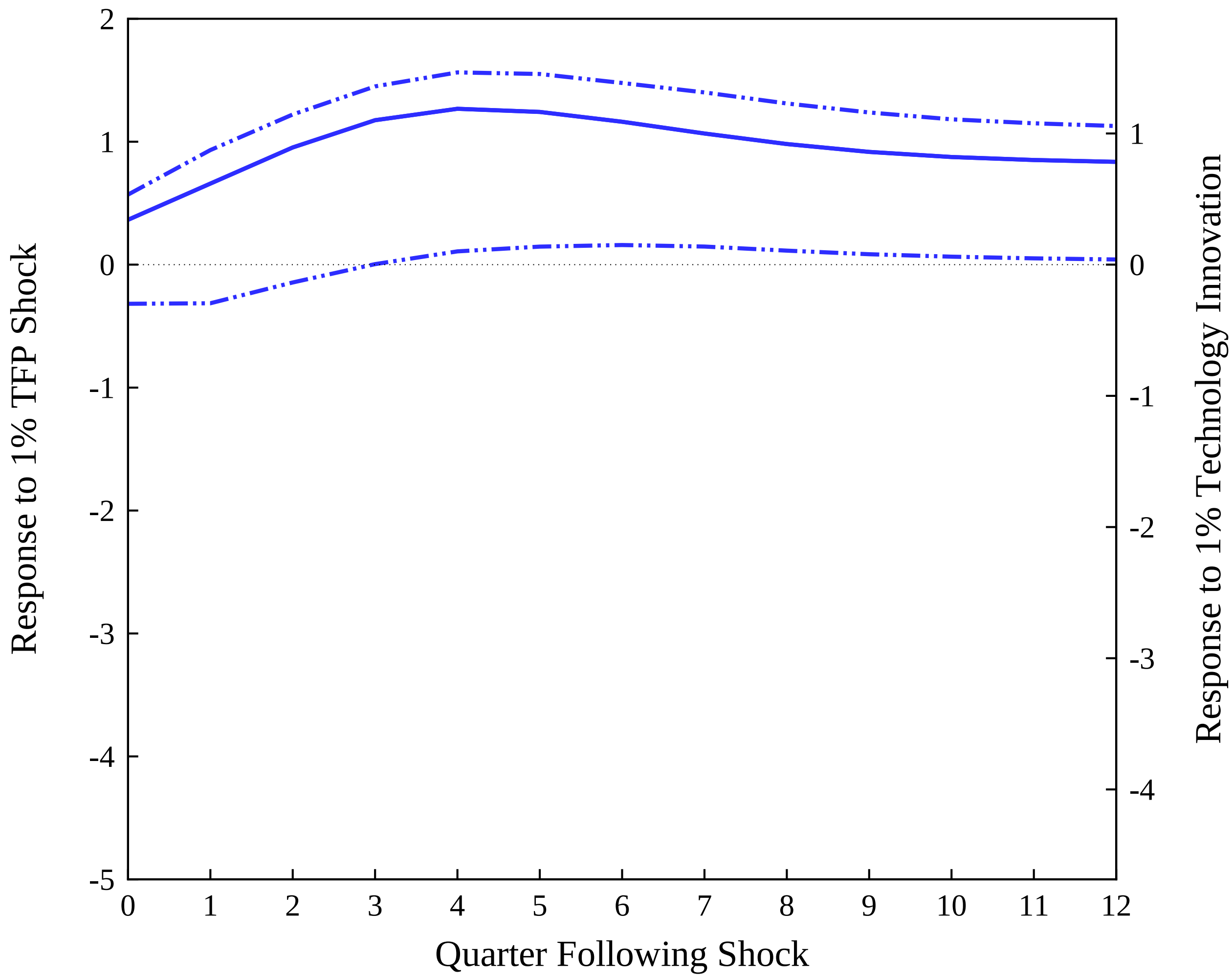


Figure 12e

Impulse Response for Galí and Rabanal (2004) Data

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 Galí and Rabanal (2004)