Technology and Business Cycles: How Well Do Standard Models Explain the Facts?

Susanto Basu*

How well do current business-cycle models explain historical output fluctuations? Almost a decade has passed since Plosser (1989) claimed that a simple real-business-cycle (RBC) model could generate simulated output with a correlation of 0.87 with actual output over the period since the Korean War. A similar observation led Prescott (1986) to claim that theory is ahead of business-cycle measurement. This paper revisits Plosser's exercise, using some recent innovations in business-cycle theory and measurement. It uses estimates of technology change recently derived by Basu, Fernald, and Kimball (1998) (henceforth BFK) and finds that a simple RBC model calibrated with these shocks produces simulated output that is *negatively* correlated with actual output. A simple dynamic general-equilibrium (DGE) model with sticky prices does somewhat better: Its impulse response to a technology improvement is qualitatively similar to that found in the data, but quantitatively the results are only moderately satisfactory. A simulation of the sticky-price model using estimates of both technology and monetary policy shocks generates model output that has a correlation of about 0.30 with actual output. Thus, current business-cycle models cannot easily explain the observed facts—measurement seems to be ahead of theory once again.

My results differ from Plosser's mainly because of the new measure of technological change that I employ. The particular series used here was

*Associate Professor of Economics, University of Michigan. The author thanks Alejandro Micco for outstanding research assistance, Craig Burnside and Michael Kiley for supplying data and programs, and discussants Mark Bils and Tom Cooley for stimulating comments. He also thanks John Fernald and Miles Kimball, who collaborated on much of the research summarized here, and provided invaluable suggestions for this paper. Finally, the author is grateful to the National Science Foundation and the Alfred P. Sloan Foundation for financial support.

estimated by BFK, but it is similar to those derived by a number of recent researchers attempting to produce better measures of short-run technical change than the standard Solow residual (for example, Burnside, Eichenbaum, and Rebelo 1996; Gali 1998). This new series shows that short-run changes in output and, especially, inputs are negatively correlated with technology improvements. As is well known, the usual RBC model produces impulse responses for output and other variables that are strongly positively correlated with technical change. Thus, it is unsurprising that the RBC model cannot duplicate the co-movements between observed variables and the new measure of technical change.

This paper makes four points. First, it reviews the evidence suggesting that technical innovation has contractionary effects in the short run. BFK argue that a major reason for the strong positive correlation of the Solow residual with output is measurement error coming from variable capital and labor utilization. Second, the paper shows that a sticky-price model with variable capital utilization can do a reasonable job of matching very short-run movements in business-cycle variables following an improvement in technology, but it also finds that the model does not capture the medium-run dynamics. Third, the paper argues that variable utilization is not just a bias; it can be an important propagation mechanism for both technology and money shocks, amplifying and propagating the effects of small disturbances.[1] However, even with this new mechanism, nominal shocks do not have persistent effects. The basic problem appears to be the standard one in the sticky-price literature, the lack of sufficiently strong propagation mechanisms (real rigidities[2]) in addition to variable utilization. Fourth, however, the paper suggests that real rigidities that are strong enough to generate substantial endogenous price stickiness in response to nominal shocks may lead to implausibly large fluctuations in response to technology shocks. Thus, producing plausible sticky-price DGE models of business cycles may be even more complicated than hitherto believed.

The first part of the paper reviews the method that BFK use to purge the Solow residual of various nontechnological components. Plosser (1989) took the standard Solow productivity residual as his measure of short-run technical change. Since then, a huge body of work has searched for other explanations for the procyclicality of productivity. This literature has advanced four main explanations for procyclical productivity. First, as Plosser assumed, procyclical productivity may reflect procyclical technology. Second, widespread imperfect competition and increasing

1 This point has also been made by Burnside and Eichenbaum (1996); Dotsey, King, and Wolman (1997); King and Rebelo (1997); and Wen (1997).

2 The term is from Ball and Romer (1990). Kimball (1995) provides an insightful discussion of the relationship between the amplification of shocks in the static setting of Ball and Romer and their propagation over time in the setting of current DGE models.

returns may lead productivity to rise whenever inputs rise. Third, as already mentioned, utilization of inputs may vary over the cycle, in a way that is not properly captured by standard input measures. Fourth, reallocation of resources across uses with different marginal products may contribute to procyclicality. For example, if different industries have different degrees of market power, then inputs will generally have different marginal products in different uses. Then aggregate productivity growth is cyclical if sectors with higher markups have input growth that is more cyclical.[3]

BFK control for the three nontechnological components of measured productivity and derive technology change as a residual. Empirically, variations in utilization and cyclical reallocation seem the most important for generating the negative correlation between technology and inputs. Given these results, it is easy to confirm that the standard RBC model, even augmented with variable capital utilization, cannot duplicate the impulse responses observed in the data. This finding may seem perplexing given the recent claim of King and Rebelo (1997) that variable capital utilization can "resuscitate" the RBC model. In the third section of this paper, I discuss why my conclusions differ from theirs.

I then partially confirm the conjecture of BFK that a sticky-price model with variable utilization can reproduce the impulse responses estimated from the data. The intuition is straightforward. As a simple example, suppose the quantity theory governs the demand for money, so output is proportional to real balances. In the short run, if the supply of money is fixed and prices cannot adjust, then real balances and hence output are also fixed. Now suppose that a positive technology shock occurs. With improved technology, firms need less labor to produce this unchanged output. As a result, they lay off workers and reduce hours. Over time, however, as prices adjust, the underlying real-business-cycle dynamics take over, and output and inputs rise.

It turns out that a more sophisticated version of this model can reproduce the initial contractionary effect of a technology improvement quite well, predicting a small fall in output and a large fall in inputs. However, the period of price stickiness is so short—and the response of the monetary authority to the initial contraction is likely to be so expansionary—that the contraction is succeeded by a boom far more quickly in the model than in the data.

Since the empirical section of the paper implies that changes in factor utilization are very important, I then investigate the extent to which variable utilization can also act as an important propagation mechanism

[3] For examples of these four explanations, see, respectively, Cooley and Prescott (1995); Hall (1988, 1990); Basu (1996), Bils and Cho (1994), and Shapiro (1996); and Basu and Fernald (1997).

in a dynamic general-equilibrium (DGE) model. One advantage of this method of controlling for utilization (taken from Basu and Kimball 1997) is that it also provides estimates of some of the critical parameters governing changes in utilization. Dotsey, King, and Wolman (1997) use the Basu-Kimball results to calibrate a sticky-price DGE model, and they claim that a model with interest-inelastic money demand, infinitely elastic labor supply, and variable capital utilization can generate persistent output fluctuations in response to nominal shocks. However, their result does not appear to extend to the case considered in this paper, where nominal interest rates are governed by a plausible Fed reaction function and where the labor supply elasticity is more realistic. In this setting, I can generate persistence only by allowing utilization to be highly variable. But the fact that I want the sticky-price model to deliver sensible impulse responses to both technology and money shocks turns out to be a binding constraint: Highly variable utilization produces more persistent dynamics in response to money shocks, but implausibly large fluctuations in response to technology shocks. Thus, extending sticky-price models to consider technology shocks is important both for matching the data, which suggest that such shocks are important, and for developing the theory of how they affect the economy.

The paper is structured as follows. The first section reviews the BFK method for estimating technology change, and the next two sections summarize the data and some of the empirical results. These results establish the facts that we want to match. The fourth section presents simple DGE models with and without nominal price rigidity and discusses their calibration. Model results are then presented, and the final section offers conclusions.

The Empirical Model

The empirical model is based only on cost-minimization by firms and one weak assumption about consumer preferences. It is thus consistent with a wide class of models including—but not limited to—the models explored in the fourth section, "A DGE Model with Variable Capital Utilization." The strategy will be to derive a series of corrected residuals, using the methods of Basu and Kimball (1997) and Basu and Fernald (1997). These residuals were first constructed, and their properties discussed, in BFK. This section of the paper reviews their methods.[4]

The Basic Setup

I assume that each firm's production function for gross output takes the following form:

[4] The presentation in this section draws on Basu and Fernald (1998).

$$Y = F(\tilde{K}, \tilde{L}, M, T) \tag{1.1}$$

The firm produces gross output, Y, using capital services $\tilde{K}$, labor services $\tilde{L}$, and intermediate inputs of materials and energy M. T indexes technology. T also includes the effects of any externalities that may exist. (For simplicity, time and firm subscripts are omitted.)

In principle, the services of labor and capital depend on both the raw quantities of these inputs (hours worked, and the capital stock) and the intensity with which they are used. Hence, labor services, $\tilde{L}$, depend on the number of employees, N, hours worked per employee, H, and the effort of each worker, E. Capital services depend on the capital stock, K, and the utilization of the capital stock, Z. Input services are therefore the following products:

$$\begin{aligned} \tilde{L} &= EHN, \\ \tilde{K} &= ZK \end{aligned} \tag{1.2}$$

I generally assume that the capital stock and the number of employees are quasi-fixed, so that firms cannot change their levels costlessly. In the short run, firms can vary their inputs of capital and labor only by varying utilization.

I assume that the firm's production function F is (locally) homogeneous of arbitrary degree γ in total inputs. Constant returns corresponds to the case where γ equals one. Formally, we can write returns to scale in two useful, and equivalent, forms. First, returns to scale equal the sum of output elasticities:

$$\gamma = \frac{F_1\tilde{K}}{Y} + \frac{F_2\tilde{L}}{Y} + \frac{F_3M}{Y}, \tag{1.3}$$

where F_J^i denotes the derivative of the production function with respect to the Jth element (that is, the marginal product of input J). Second, once we assume that firms minimize cost, we can denote the firm's cost function by $C(Y)$. (In general, the cost function also depends on the prices of the variable inputs and the quantities of any quasi-fixed inputs, although for simplicity I suppress those terms here.) The *local* degree of returns to scale equals the inverse of the elasticity of cost with respect to output (see Varian 1984, p. 68):

$$\gamma(Y) = \frac{C(Y)}{YC'(Y)} = \frac{C(Y)/Y}{C'(Y)} = \frac{AC}{MC}, \tag{1.4}$$

where AC equals average cost, and MC equals marginal cost. Note that increasing returns, for example, may reflect overhead costs or decreasing marginal cost; both imply that average cost exceeds marginal cost. If increasing returns take the form of overhead costs, then $\gamma(Y)$ is not a constant structural parameter, but depends on the level of output the firm produces. As production increases, returns to scale fall as the firm moves down its average cost curve.

As equation (1.4) shows, there is no necessary relationship between the degree of returns to scale and the slope of the marginal cost curve. Indeed, increasing returns is compatible with increasing marginal costs, as in the standard Chamberlinian model of imperfect competition. One can calibrate the slope of the marginal cost curve from the degree of returns to scale only by assuming no fixed costs. This point is an important one, because it is the slope of the marginal cost curve that determines the slopes of the factor demand functions, which in turn are critical for determining the results of DGE models, like the one in the fourth section, below. A number of studies have used estimates of the degree of returns to scale to calibrate the slope of marginal cost: This procedure is not legitimate.

Firms may charge a price P that is a markup, μ, over marginal cost. That is, $\mu = P/MC$. Returns to scale γ is a technical property of the production function, while the markup μ is essentially a behavioral parameter, depending on the firm's pricing decision. However, the following identity links the two parameters:

$$\gamma = \frac{C(Y)}{YC'(Y)} = \frac{P}{C'(Y)} \frac{C(Y)}{PY} = \mu(1 - s_\pi), \tag{1.5}$$

where s_π is the share of pure economic profit in gross revenue. As long as pure economic profits are small (Rotemberg and Woodford 1995 provide a variety of evidence suggesting that profit rates are close to zero), equation (1.5) shows that μ approximately equals γ. Large markups, for example, require large increasing returns.

Given low estimated profits, equation (1.5) also shows that strongly diminishing returns (γ less than one) imply that firms consistently price output below marginal cost (μ less than one). Since pricing below marginal cost makes no economic sense, I conclude that firm-level returns to scale must either be constant or increasing. Note also that increasing returns *requires* that firms charge a markup, as long as firms do not make losses.

The Solow-Hall Approach

Solow's (1957) seminal contribution involves differentiating the production function and using the firm's first-order conditions for cost minimization. Solow assumed constant returns to scale and perfect competition, so the first-order conditions (discussed below) imply that output elasticities are observed in the data as factor shares in revenue. Hall (1988, 1990) builds on Solow's contribution, extending it to the case of increasing returns and imperfect competition. Under these conditions, output elasticities are not observed, since neither returns to scale nor markups are observed. However, Hall derives a simple regression equation, which he then estimates. This section extends Hall's approach by using gross-output data and taking account of variable factor utilization.

Taking the logarithm of the production function (1.1), and differentiating it totally, one gets

$$dy = \frac{F_1 ZK}{Y}(dk + dz) + \frac{F_2 EHN}{Y}(de + dh + dn) + \frac{F_3 M}{Y} dm + dt, \tag{1.6}$$

where lower-case letters represent logs. Without loss of generality, I have normalized to one the elasticity of output with respect to technology.

Suppose firms take the price of all J inputs, P_J, as given. They may have market power in output markets. If all factors are freely variable, then the first-order conditions for cost-minimization imply that:

$$PF_J = \mu P_J. \tag{1.7}$$

In other words, firms set the value of a factor's marginal product equal to a markup over the factor's input price. Equivalently, rearranging the equation by dividing through by μ, this condition says that firms equate each factor's marginal revenue product ($(P/\mu)F_J$) to the factor's price.

Equation (1.7) still holds in the case where some factors are quasi-fixed, as long as we define the input price of the quasi-fixed factors as the appropriate *shadow* price, or implicit rental rate. I return to this point in a later subsection, when I specify a more complicated dynamic cost-minimization problem. Note also that the price of capital, P_K, must be defined as the *rental price* (or shadow rental price) of capital. In particular, if the firm makes pure economic profits, these are generally paid to capital: These profits must be subtracted before computing the rental price. (Note that these profits are over and above the quasi-rents that can accrue to a fixed factor, which are incorporated into the rental price of capital.)

Using equation (1.7), we can write each output elasticity as the product of the markup multiplied by total expenditure on each input divided by total revenue. Thus, for example,

$$\frac{F_1 ZK}{Y} = \mu \frac{P_K K}{PY} \equiv \mu s_K. \tag{1.8}$$

(Note that the marginal product of capital is $F_1 Z$, since the services from a machine depend on the rate at which it is being utilized.)

Substituting these expressions for the output elasticities into (1.6), we get the basic estimating equation for the markup:

$$\begin{aligned} dy &= \mu[s_K(dk + dz) + s_L(dn + dh + de) + s_M dm] + dt \\ &= \mu[s_K dk + s_L(dn + dh) + s_M dm] + \mu[s_K dz + s_L de] + dt \\ &\equiv \mu dx + \mu du + dt, \end{aligned} \tag{1.9}$$

where dx is a share-weighted average of conventional (observed) input growth, and du is a weighted average of unobserved variation in utilization and effort. Note that the shares are the total *cost* of each type of input divided by total *revenue*. Thus, the shares in dx sum to less than one if firms make pure profits.

The derivation so far is in the spirit of Hall (1990), generalized to include variable utilization. Hall, in turn, generalizes Solow (1957) to the case of imperfect competition. (Both Hall and Solow considered variable utilization, at least in principle.) Solow's derivation assumes perfect competition and constant returns, so μ equals one. Since there are no economic profits in that world, as shown by equation (1.5), capital's share can be taken as a residual.

Note that using equation (1.5), we can rewrite equation (1.9) in terms of returns to scale γ. In this case, the weights used to calculate weighted-average inputs dx are cost shares, which sum to one. Hall (1990) pioneered this latter approach, although no economic difference is found between thinking of the output elasticity of inputs in terms of the markup and thinking in terms of returns to scale, and the data requirements are the same in the two cases.

It is important to note that the derivation relies solely on cost minimization: Profit maximization is irrelevant. This is a large advantage, since we can ignore the firm's behavior in product markets, which may be very complex. For example, firms may sell output with sticky prices (as in the model in the fourth section, below), or engage in strategic interactions in a repeated-game setting, but the existence of such behavior does not affect the results.

Several practical issues need to be resolved before estimating equation (1.9). First, we must figure out the appropriate prices to use in calculating weights. With quasi-fixed inputs, the appropriate shadow price is not, in general, the observed factor price. Second, we must find suitable proxies for du. To address these practical issues, we next specify a cost-minimization problem that provides a framework for analysis.

A Dynamic Cost-Minimization Problem

Although the problem is relatively complicated, specifying a particular dynamic cost-minimization problem provides insight into several practical issues in attempting to estimate equation (1.9). In the subsection "Variable Utilization," below, it also provides proxies for unobserved utilization, as well as a method for estimating crucial parameters used to calibrate the models of the fourth section.

The firm is modeled as facing adjustment costs in both investment and hiring, so that both the amount of capital (number of machines and buildings), K, and employment (number of workers), N, are quasi-fixed. I model quasi-fixity for two reasons. First, I want to examine the effect of quasi-fixity per se on estimates of production-function parameters and firm behavior. Second, quasi-fixity is necessary for a meaningful model of variable factor utilization. Higher utilization must be more costly to the firm, otherwise factors would always be fully utilized. If increasing the rate of investment or hiring had no cost, firms would always keep utilization at its minimum level and vary inputs using only the extensive margin, hiring and firing workers and capital costlessly. Only if it is costly to adjust along the extensive margin is it sensible to adjust along the intensive margin, and pay the costs of higher utilization.[5]

While capital and labor have adjustment costs, I assume that the number of hours per week for each worker, H, can vary freely, with no adjustment cost. In addition, both capital and labor have freely variable utilization rates. For both capital and labor, the benefit of higher utilization is its multiplication of effective inputs. I assume two costs of increasing capital utilization, Z. First, capital depreciates faster because of extra wear and tear. Second, firms may have to pay a shift premium to compensate employees for working at night or at other undesirable times. I take Z to be a continuous variable for simplicity, although variations in the workday of capital (that is, the number of shifts) are perhaps the most plausible reason for variations in utilization. The variable-shifts model has had considerable empirical success in manufacturing data, where, for a short period of time, one can observe the number of shifts directly.[6] The cost of higher labor utilization, E, is a higher disutility on the part of

[5] One does not require *internal* adjustment costs to model variable factor utilization in an aggregative model (see, for example, Burnside and Eichenbaum 1996), since changes in input demand on the part of the representative firm change the aggregate real wage and interest rate, so in effect the concavity of the representative consumer's utility function acts as an adjustment cost that is *external* to the firm. However, if one wants to model the behavior of firms that vary utilization in response to idiosyncratic changes in technology or demand—obviously the case in the real world—then one is forced to posit the existence of internal adjustment costs in order to have a coherent model of variable factor utilization. (Both of these observations are found in Haavelmo's (1960) treatment of investment.)

[6] See, for example, Shapiro (1996).

workers that must be compensated with a higher wage. I allow for the possibility that this wage is unobserved from period to period, as might be the case if wage payments are governed by an implicit contract in a long-term relationship.

Consider the following cost-minimization problem for the representative firm of an industry:

$$\min_{Z,E,H,M,I,A} C(Y) = \int_0^\infty e^{-\int_0^s r d\tau} [WNG(H,E) + P_M M + WN\Psi(A/N) + P_I KJ(I/K)]\, ds \tag{1.10}$$

subject to

$$Y = F(ZK, EHN, M, T) \tag{1.11}$$

$$\dot{K} = I - \delta(Z)K \tag{1.12}$$

$$\dot{N} = A. \tag{1.13}$$

The production function and inputs are as before. In addition, I is gross investment, and A is hiring net of separations. $WG(H,E)$ is total compensation per worker, where W is the base wage (compensation may take the form of an implicit contract, and hence not be observed period-by-period); $WN\Psi(A/N)$ is the total cost of changing the number of employees; $P_I KJ(I/K)$ is the total cost of investment; P_M is the price of materials. $\delta(Z)$ is the variable rate of depreciation. I continue to omit time subscripts for clarity.

Using a perfect-foresight model amounts to making a certainty-equivalence approximation. But even departures from certainty equivalence should not disturb the key results, which rely only on intratemporal optimization conditions rather than intertemporal ones.

I assume that Ψ, J, and δ are convex, and make the appropriate technical assumptions on G in the spirit of convexity and normality.[7] It is also helpful to make some normalizations in relation to the normal or "steady-state" levels of the variables. Using an asterisk to denote these steady-state levels, let $\delta(Z^*) = \delta^*$, $J(\delta^*) = 0$, $J'(0) = 1$, $\Psi(0) = 0$. I also assume that the marginal employment adjustment cost is zero at a constant level of employment: $\Psi'(0) = 0$.

I solve the representative firm's problem using the standard current-

[7] The conditions on G are easiest to state in terms of the function Φ defined by $\ln G(H, E) = \Phi(\ln H, \ln E)$. Convex Φ guarantees a global optimum; assuming $\Phi_{11} > \Phi_{12}$ and $\Phi_{22} > \Phi_{12}$ ensures that optimal H and E move together.

value Hamiltonian, letting λ, q, and θ be the multipliers on constraints (1.11), (1.12), and (1.13) respectively. Using numerical subscripts for derivatives of the production function F with respect to its first, second, and third arguments, and literal subscripts for derivatives of the labor cost function G, the firm's six intratemporal first-order conditions for cost-minimization are:

$$\text{Z:} \qquad \lambda K\, F_1(ZK, EHN, M; T) = qK\delta'(Z) \tag{1.14}$$

$$\text{H:} \qquad \lambda EN\, F_2(ZK, EHN, M; T) = WN\, G_H(H, E) \tag{1.15}$$

$$\text{E:} \qquad \lambda HN\, F_2(ZK, EHN, M; T) = WN\, G_E(H, E) \tag{1.16}$$

$$\text{M:} \qquad \lambda F_3(ZK, EHN, M; T) = P_M \tag{1.17}$$

$$\text{A:} \qquad \theta = W\Psi'(A/N) \tag{1.18}$$

$$\text{I:} \qquad q = P_I J'(I/K). \tag{1.19}$$

The Euler equations for the capital stock and employment are:

$$\dot{q} = [r + \delta(Z)]q - \lambda ZF_1 + P_I[J(I/K) - (I/K)J'(I/K)] \tag{1.20}$$

$$\dot{\theta} = r\theta - \lambda EHF_2 + WG(H, E) + W[\Psi(A/N) - (A/N)\Psi'(A/N)]. \tag{1.21}$$

As the Lagrange multiplier associated with the level of output, λ can be interpreted as marginal cost. Since the firm internally values output at marginal cost, λF_1 is the marginal value product of effective capital input, λF_2 is the marginal value product of effective labor input, λF_3 is the marginal value product of materials input, and λF_4 is the marginal value product of energy input.[8] Using the definition that the markup, μ, equals the ratio of output price, P, to marginal cost, I rewrite λ as:

$$\lambda = C'(Y) = \frac{P}{\mu}. \tag{1.22}$$

Note that equation (1.22) is just a definition, not a theory determining the markup. The markup depends on the solution of the firm's more complex profit-maximization problem, which we do not need to specify at all.

Equations (1.20) and (1.21) implicitly define the shadow (rental) prices of labor and capital:

---

[8] For the standard static profit-maximization problem, of course, marginal cost equals marginal revenue, so these are also the marginal revenue products.

$$\lambda Z F_1 = [r + \delta(Z)]q - \dot{q} + P_I[J(I/K) - (I/K)J'(I/K)] \equiv P_K \tag{1.20$'$}$$

$$\lambda E H F_2 = r\theta - \dot{\theta} + WG(H, E) + W[\Psi(A/N) - (A/N)\Psi'(A/N)] \equiv P_L \tag{1.21$'$}$$

As usual, the firm equates the marginal value product of each input to its shadow price. Note that with these definitions of shadow prices, the atemporal first-order condition (1.7) is satisfied for all inputs. For some intuition, note that equation (1.20′) is the standard first-order equation from a q-model of investment. In the absence of adjustment costs, the value of installed capital q equals the price of investment goods P_I, and the "price" of capital input is then just the standard Hall-Jorgenson rental cost of capital, $(r + \delta)P_I$. With investment adjustment costs, there is potentially an extra return to owning capital, through capital gains $\dot{q}$ (as well as extra terms that reflect the fact that investing today incurs additional adjustment costs, but produces the benefit of lowering adjustment costs in the future).

The intuition for labor in equation (1.21′) is similar. Consider the case where labor can be adjusted freely, so that it is *not* quasi-fixed. Then adjustment costs ψ are always zero; so is the multiplier θ, since constraint (1.13) does not bind. In this case, as we expect, (1.21′) says that the shadow price of labor input to the firm—the right side of (1.21′)—just equals the (effort-adjusted) compensation $WG(H,E)$ received by the worker. Otherwise, the quasi-fixity implies that the shadow price of labor to a firm may differ from the compensation received by the worker.

Implementation in Discrete Time

I now turn to issues of estimation. Equations (1.6) and (1.9) hold exactly in continuous time, if the values of the output elasticities are adjusted continuously. In discrete time, if the elasticities are treated as time-invariant, then equation (1.6) is a first-order approximation (in logs) to any general production function. For a consistent first-order approximation, one should then treat equation (1.9) as representing small deviations from a steady-state growth path and evaluate derivatives of the production function at the steady-state values of the variables. Thus, to calculate the shares in equation (1.9), one should use steady-state prices and quantities and, hence, treat the shares as constant over time. The markup is then also taken as constant.

For example, in the first-order approach, we want the steady-state output elasticity for capital, up to the unknown scalar μ. Using asterisks to denote steady-state values, we use equations (1.19), (1.20′), and the normalizations to compute the steady-state output elasticity of capital:

$$\frac{F_1^* U^* K^*}{Y^*} = \mu^* \frac{P_K^* K^*}{P^* Y^*} \equiv \mu^* \frac{(r^* + \delta^*) P_I^* K^*}{P^* Y^*}. \tag{1.23}$$

Note that the steady-state user cost of capital is the frictionless Hall-Jorgenson (1967) rental price.[9] Since quasi-fixity matters only for the adjustment to the steady state, in the steady state $q = P_I$ and $\dot{q} = 0$. Operationally, I calculate the Hall-Jorgenson user cost for each period and take the time average of the resulting shares as an approximation to the steady-state share. I proceed analogously for the other inputs. In the final estimating equation for (1.9), I use logarithmic differences in place of output and input growth rates, and use steady-state shares for the weights.

Thus, I can construct the index of observable inputs, dx, and take the unknown μ^*, multiplying it as a parameter to be estimated. We can use a variety of approaches to control for the unobserved du; some of them are discussed in the next section. In any case, we have to use instruments that are orthogonal to the technology shock dt, since technology change is generally contemporaneously correlated with input use (observed or unobserved).[10]

Variable Utilization

Before we can estimate μ from equation (1.9), we need to settle on a method for dealing with changes in utilization, du. A priori reasoning—and comparisons between results that control for du and those that do not—argue that du is most likely positively correlated with dx; thus, ignoring it leads to an upward-biased estimate of μ^*. Three general methods have been proposed. First, one can try to observe du directly using, say, data on shift work. When possible this option is clearly the preferred one, but data availability often precludes its use.[11] Second, one can impose a priori restrictions on the production function.[12] Third, one

[9] In practice, one would also include various tax adjustments. We do so in the empirical work but omit them in the model to keep the exposition simple.

[10] Olley and Pakes (1996) propose an insightful alternative to the usual instrumental-variables estimation strategy; see Griliches and Mairesse (1995) for an excellent discussion. However, their procedure generally cannot be used when estimating structural parameters governing changes in utilization, because it relies on using investment as a proxy for changes in technology, dt. The method discussed in the next subsection uses investment as a proxy for the shadow value of installed capital; thus, we cannot use the Olley-Pakes procedure and identify all the structural parameters of the model.

[11] In the United States, shift-work data are available solely for manufacturing industries, and then only for a few years. The only data set on worker effort that we know of is the survey of British manufacturing firms used by Schor (1987).

[12] For example, Jorgenson and Griliches (1967) assume that the unobserved service flow of capital is proportional to electricity use. Burnside, Eichenbaum, and Rebelo (1996) have recently used this assumption to derive utilization-adjusted estimates of technology shocks.

can derive links between the unobserved du and observable variables using first-order conditions like equations (1.14) to (1.19). Both the second and third approaches imply links between the unobserved du and observable variables, which can be used to control for changes in utilization.

Bils and Cho (1994), Burnside and Eichenbaum (1996), and Basu and Kimball (1997) argue that one can also control for variable utilization using the relationships between observed and unobserved variables implied by first-order conditions like equations (1.14) to (1.19). The discussion here follows Basu and Kimball.

They begin by assuming a generalized Cobb-Douglas production function:

$$F(ZK, EHN, M; Z) = Z\Gamma((ZK)^{\alpha_K}(EHN)^{\alpha_L}M^{\alpha_M}), \tag{1.24}$$

where Γ is a monotonically increasing function. In their case this assumption is not merely a first-order approximation, because they make use of the second-order properties of equation (1.24), particularly the fact that the output elasticities are constant. Although they argue that one can relax the Cobb-Douglas assumption, I shall maintain it throughout the discussion.

Equations (1.15) and (1.16) can be combined into an equation implicitly relating E and H:

$$\frac{HG_H(H,E)}{G(H,E)} = \frac{EG_E(H,E)}{G(H,E)}. \tag{1.25}$$

The elasticity of labor costs with respect to H and E must be equal, because on the benefit side the elasticities of effective labor input with respect to H and E are equal. Given the assumptions on G, (1.25) implies a unique, upward-sloping E–H expansion path, so that we can write

$$E = E(H), \qquad E'(H) > 0. \tag{1.26}$$

Equation (1.26) says that the unobservable intensity of labor utilization E can be expressed as a monotonically increasing function of the observed number of hours per worker, H.

Finding the marginal product of capital from (1.14), substituting into (1.8), and rearranging, we find that the level of capital utilization depends on the degree to which the current marginal value product of capital exceeds future marginal products:

$$Z\delta'(Z) = \lambda\gamma\alpha_K \frac{Y}{qK}. \tag{1.27}$$

Since fluctuations in marginal cost λ, returns to scale γ, and the marginal value of capital q are difficult to observe directly, we would like to express these factors in terms of other variables that are more readily observed.

The problem with trying to measure q directly is not just the difference between the marginal and average value of capital but also the noisiness of the asset prices one would use to gauge the average value of capital. Instead of trying to measure q directly, Basu and Kimball use equation (1.19) to express q as the price of investment goods times a function of I/K. (Note that Tobin's q is actually q/P_I in my notation.) Equation (1.19) can be inverted to say that I/K is a function of Tobin's q.

The first-order condition for materials usage (1.17) is the key to expressing the product $\lambda\gamma$ in terms of observables. Combining this equation with the expression for the marginal products of materials, we find

$$\lambda\gamma = \frac{P_M M}{\alpha_M Y}. \tag{1.28}$$

Thus, we can measure the marginal value product of capital as:

$$\lambda\gamma\alpha_K = \frac{\alpha_K}{\alpha_M}\frac{P_M M}{Y}. \tag{1.29}$$

Substituting the expression for the marginal revenue product of capital (equation 1.29) and the expression for q (equation (1.19) into (1.27) leads to the desired expression for capital utilization in terms of observed variables and the ratio α_K/α_M:

$$Z\delta'(Z) = \frac{\alpha_K}{\alpha_M}\frac{P_M M}{P_I K}\frac{1}{J'(I/K)}. \tag{1.30}$$

Define a number of elasticities in terms of steady-state values of different variables; let

$$\zeta \equiv \frac{H^* E'(H^*)}{E(H^*)},$$

$$\Delta \equiv \frac{Z^*\delta''(Z^*)}{\delta'(Z^*)},$$

and

$$j \equiv \frac{(I/K)^* J''((I/K)^*)}{J'((I/K)^*)} = \frac{\delta^* J''(\delta^*)}{J'(\delta^*)}.$$

Thus, from equation (1.26),

$$d \ln(EHN) = dn + dh + de = dn + (1 + \zeta)dh. \qquad (1.31)$$

With a constant α_K/α_M,[13] (1.31) implies

$$dz = \frac{1}{1 + \Delta}(dp_M + dm - dp_I - dk) - \frac{j}{1 + \Delta}(di - dk). \qquad (1.32)$$

Putting everything together, we have an estimating equation that controls for variable utilization:

$$dy = \mu^* dx + \mu^* \zeta s_L dh + \frac{\mu^*}{1 + \Delta} s_K(dp_M + dm - dp_I - dk) - \frac{\mu^* j}{1 + \Delta} s_K(di - dk) + dt. \qquad (1.33)$$

This specification controls for both labor and capital utilization, without making special assumptions about separability or homotheticity. However, for this simple derivation, the Cobb-Douglas functional form is important. One payoff of the Basu-Kimball approach is that it allows one not only to control for variations in utilization, du, but also to estimate the key elasticities governing changes in Z and E, which will be used in the model simulations below.[14] The residual from this equation is a measure of technology change, cleansed of distortions coming from imperfect competition and variable utilization.

The Definition of Technology Change

How are the firm-level technology shocks defined (implicitly) by equation (1.33), related to aggregate technology shocks? Aggregate technology change is sometimes defined from a macro (top down) perspective, and sometimes from a micro (bottom up) perspective. A sensible macro definition is the change in final output (that is, $C + I +$

[13] This equation is where the Cobb-Douglas assumption matters; Basu and Kimball differentiate (1.31) assuming that α_K/α_M is a constant. Their theory allows for the fully general case where the ratio of the elasticities is a function of all four input quantities, but they argue that pursuing this approach would demand too much of the data and instruments.

[14] So far, we have abstracted from the existence of a shift premium. However, utilizing capital more intensively by running it for extra shifts may require paying workers on later shifts a higher base wage to compensate for the disutility of working at non-standard hours. Basu and Kimball extend the model above to incorporate a shift premium. They show that an estimating equation with the same three extra variables as (1.33) controls for utilization even in this extended model. Thus, the technology residuals and markup estimates are correct even in this more general framework. However, the parameters governing changes in utilization—particularly Δ—are no longer identified once the model is generalized to include a shift premium.

$G + X - M$), for given aggregate primary inputs. A sensible micro definition is an appropriately weighted average of firm-level technology change. With constant returns and perfect competition, these two perspectives are equivalent (Domar 1961; Hulten 1978). Rotemberg and Woodford (1995) show that equivalence also holds with imperfectly competitive product markets, under certain restrictive conditions: perfect factor markets, and all firms having identical separable gross-output production functions, charging prices that are the same markup over marginal cost, and always using intermediate inputs in fixed proportions to gross output.

If the Rotemberg-Woodford assumptions fail—if, for example, factor markets are imperfectly competitive or firms have different degrees of market power—then the two perspectives lead to different definitions; that is, aggregate technology from a macro perspective is not a weighted average of firm-level technology.[15] For example, suppose differences in markups or factor payments across firms lead the same factor to have a different social value for its marginal product in different uses. Then changes in the *distribution* of inputs can affect final output, even if firm-level technology and aggregate inputs are held constant. Conceptually, however, we may not want to count such variation as "technology change," since it can occur with no change in the technology available to any firm.

Now consider the following definition of technical change: the increase in aggregate output, holding fixed not only aggregate primary inputs, but also their distribution across firms and the materials/output ratio at each firm. Although this definition is close in spirit to the macro perspective, it also corresponds to a reasonable micro definition, since aggregate technology changes only if firm-level technology changes. Indexing firms by i, Basu and Fernald (1997) show that this measure of technical change equals:

$$dt = \sum_i w_i \frac{dt_i}{1 - \mu_i s_{Mi}}, \tag{1.34}$$

where w_i is the firm's share of aggregate nominal value added:

$$w_i = \frac{P_i Y_i - P_{Mi} M_i}{\sum_i (P_i Y_i - P_{Mi} M_i)} \equiv \frac{P_i^V V_i}{P^V V}.$$

Conceptually, this measure first converts the gross-output technology shocks to a value-added basis by dividing through by $1 - \mu s_M$. (A

[15] Basu and Fernald (1997).

value-added basis is desirable because of the national accounts identity, which tells us that aggregate final expenditure equals aggregate value added.)[16] These value-added shocks are then weighted by the firm's share of aggregate value added.

Equation (1.34) defines a "micro" measure of technical change, since it changes only if firm-level production technology changes. However, it also nests the Rotemberg-Woodford definition of technology as a special case, and thus it correctly measures "macro" technical change under their conditions. This property is desirable, since the Rotemberg-Woodford assumptions are implicit or explicit in most dynamic general-equilibrium models with imperfect competition. I thus focus on definition (1.34) in constructing the aggregate technology series.

However, the measure defined in equation (1.34) has the disadvantage that it requires one to know (or estimate) the firm-level markups. Domar (1961) and Hulten (1978) propose a different definition of aggregate technology:

$$dt' = \sum_i w_i \frac{dt_i}{1 - s_{Mi}}. \tag{1.35}$$

They show that equation (1.35) satisfies both the micro and macro definitions of technical change when there are constant returns and perfect competition: Note that (1.34) reduces to (1.35) when μ equals one everywhere.

With imperfect competition, the Domar-weighted measure shows how much increases in firm-level technical change increase final output, holding fixed both the aggregate quantities and the distributions of primary *and* intermediate inputs. This definition is unappealing, since it corresponds to a thought experiment where firms are not allowed to use more intermediate inputs even when they receive favorable technology shocks. However, it does have the advantage that it does not require knowledge of sectoral markups. BFK thus also use this measure of technical change to check the robustness of the primary measure, and they find that their results are unaffected by using one measure rather than the other.

We define changes in aggregate utilization as the contribution to

[16] Basu and Fernald (1997) discuss this conversion to value added at length. To understand why $(1 - \mu s_M)$ is the right denominator, consider the case where a firm uses materials in fixed proportion to output, and receives a gross-output technology innovation dt. The firm's output (which, for simplicity, we can assume is sold only for final demand) increases both because of the technology improvement and because of the productive contribution of the required additional materials. Since the marginal product of materials is μs_M, output increases by $dy = dt + \mu s_M dm$. Since $dm = dy$ (the materials/output ratio is fixed), this equation implies that the change in output is $dt/(1 - \mu s_M)$.

final output of changes in firm-level utilization. This, in turn, is a weighted average of firm-level utilization change du_i:

$$du = \sum_i w_i \frac{\mu_i du_i}{1 - \mu_i s_{Mi}} \tag{1.36}$$

Note from equation (1.9) that $\mu_i du_i$ enters in a manner parallel to dt_i and hence (1.36) parallels (1.34).

Data and Method

The Data

I now construct a measure of "true" aggregate technology change, dt, and explore its properties. As discussed in the previous section, I estimate technology change at a disaggregated level, and then aggregate. The aggregate is the private U.S. economy, and the "firms" are 34 industries; for manufacturing, these industries correspond roughly to the 2-digit SIC level.

Each industry contains thousands or tens of thousands of firms, so it may seem odd to take industries as firms. Unfortunately, no firm-level data sets span the economy. In principle, I could focus on a subset of the economy, using the Longitudinal Research Database, say; however, narrowing the focus requires sacrificing a macroeconomic perspective, as well as panel length and data quality. By focusing on aggregates, the paper complements existing work that uses small subsets of the economy.

I use data compiled by Dale Jorgenson and Barbara Fraumeni on industry-level inputs and outputs. These data consist of a panel of 33 private industries (including 21 manufacturing industries) that cover the entire U.S. nonfarm private economy. These sectoral accounts seek to provide accounts that are, to the extent possible, consistent with the economic theory of production. Output is measured as gross output, and inputs are separated into capital, labor, energy, and materials. These data are available from 1947 to 1989; in the empirical work, however, I restrict my sample to 1950 to 1989, since the money shock instrument is not available for previous years. For a complete description of the data set, see Jorgenson, Gollop, and Fraumeni (1987).

I compute capital's share s_K for each industry by constructing a series for required payments to capital. I follow Hall and Jorgenson (1967) and Hall (1990), and estimate the user cost of capital R. For any type of capital, the required payment is then $RP_K K$, where $P_K K$ is the current-dollar value of the stock of this type of capital. In each sector, I use data on the current value of the 51 types of capital, plus land and inventories, distinguished by the U.S. Bureau of Economic Analysis in constructing the national product accounts. Hence, for each of these 53 assets, indexed by s, the user cost of capital is

$$R_s = (r + \delta_s)\,\frac{(1 - ITC_s - \tau d_s)}{(1 - \tau)},\; s = 1 \text{ to } 53. \tag{2.1}$$

r is the required rate of return on capital (and on all other assets except money), and δ_s is the depreciation rate for assets of type s. ITC_s is the asset-specific investment tax credit, τ is the corporate tax rate, and d_s is the asset-specific present value of depreciation allowances. I follow Hall (1990) in assuming that the required return r equals the dividend yield on the S&P 500. Jorgenson and Yun (1991) provide data on ITC_s and d_s for each type of capital good. Given required payments to capital, computing s_K is straightforward.

For the empirical work, we need instruments that are uncorrelated with technology change. I use two of the Hall-Ramey instruments: the growth rate of the price of oil deflated by the GDP deflator and the growth rate of real government defense spending.[17] (I use the contemporaneous value and one lag of each instrument.) To these I add a version of the instruments used by Burnside (1996), quarterly Federal Reserve "policy shocks" from an identified VAR. I use the sum of the four quarterly policy shocks in year $t - 1$ as instruments for input growth in year t.[18]

Estimating Technology Change

To estimate "firm-level" technology change, I estimate equation (1.33) for each industry. Although I could estimate these equations separately for each industry (and indeed do so as a check on results), some parameters, particularly the utilization proxies, are then estimated

[17] We drop the third instrument, the political party of the President, because it appears to have little relevance in any industry. Burnside (1996) shows that the oil price instrument is generally quite relevant, and defense spending explains a sizable fraction of input changes in the durable-goods industries.

[18] The qualitative features of the results in the next section, "Empirical Results," appear robust to using different combinations and lags of the instruments. On a priori grounds, the set I choose seems preferable to alternatives—all of the variables have strong grounds for being included. In addition, the set chosen has the best overall fit (measured by mean and median F statistic) of the a priori plausible combinations considered. Of course, Hall, Rudebusch, and Wilcox (1996) argue that with weak instruments, one does not necessarily want to choose the instruments that happen to fit best in sample; for example, if the "true" relevance of all the instruments is equal, the ones that by chance fit best in sample are in fact those with the largest small sample bias. That case is probably not a major concern here, since the instrument set we choose fits well for all industry groupings; for example, it is the one we would choose based on a rule of, say, using the instruments that fit best in durables industries as instruments for nondurables industries, and vice versa.

rather imprecisely. To mitigate this problem, I combine industries into four groups, estimating equations that restrict the utilization parameters to be constant within industry groups. Thus, for each group we have

$$dy_i = c_i + \mu_i dx_i + adh_i + b(dp_{Mi} + dm_i - dp_{Ii} - dk_i) + c(di_i - dk_i) + dt_i. \tag{2.2}$$

The markup μ_i differs by industries within a group (Burnside (1996) emphasizes the importance of allowing this heterogeneity). The groups are durables manufacturing (11 industries); nondurables manufacturing (10); natural-resource extraction, such as mining and petroleum extraction (4); and all others, mainly services and utilities (8). To avoid the "transmission problem" of correlation between technology shocks and input use, I estimate each system using Three-Stage Least Squares, using the instruments noted above.

After estimating equation (2.2), the sum of the industry-specific constant $\hat{c}_i$ and residual $d\hat{t}_i$ measures technology change in the gross-output production function. Since I am ultimately interested in the aggregate effects of technology shocks, I take an appropriately weighted average of the firm-level estimates of technology change, using equation (1.34).

Empirical Results

This section summarizes the properties of the "true" technology series; the results are taken from BFK. These results serve two purposes. First, they explain the properties of the technology series, which will be used as an input into the model simulations of the fifth section, "Simulation Results," below. Second, they allow us to compute impulse responses to technology improvements, which will serve as benchmarks for assessing the performance of the models.

Basic Correlations

Table 1 reports summary statistics for three series: (i) the Solow residual; (ii) a series that makes no utilization corrections, but corrects only for aggregation biases; and (iii) a "technology" measure based on equation (2.2). Note that the first measure uses aggregate data alone, whereas the other two are based on sectoral regression residuals, which are aggregated using equation (1.4).

The corrected series have about the same mean as the Solow residual. However, the variance is substantially smaller: The variance of the fully corrected series is less than one-third that of the Solow residual, so the standard deviation (shown in the second column) is only about 55 percent as large. The reported minimums show negative technical change in some periods, but the lower variance of the technology series implies

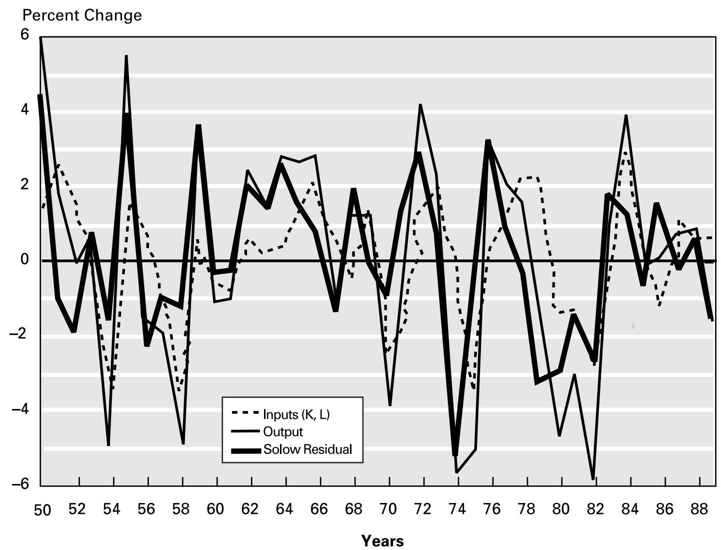
Table 1
Descriptive Statistics for Technology Residuals

	Mean	Standard Deviation	Minimum	Maximum
Solow Residual	.011	.022	−.044	.066
Technology Residual (No Utilization Correction)	.012	.016	−.034	.050
Technology Residual (Full Basu-Kimball (1997) Correction)	.013	.012	−.013	.032

that the probability of negative estimates is much lower. For example, the Solow residual is negative in 12 out of 40 years; the fully corrected residual is negative in only 5 out of 40 years.

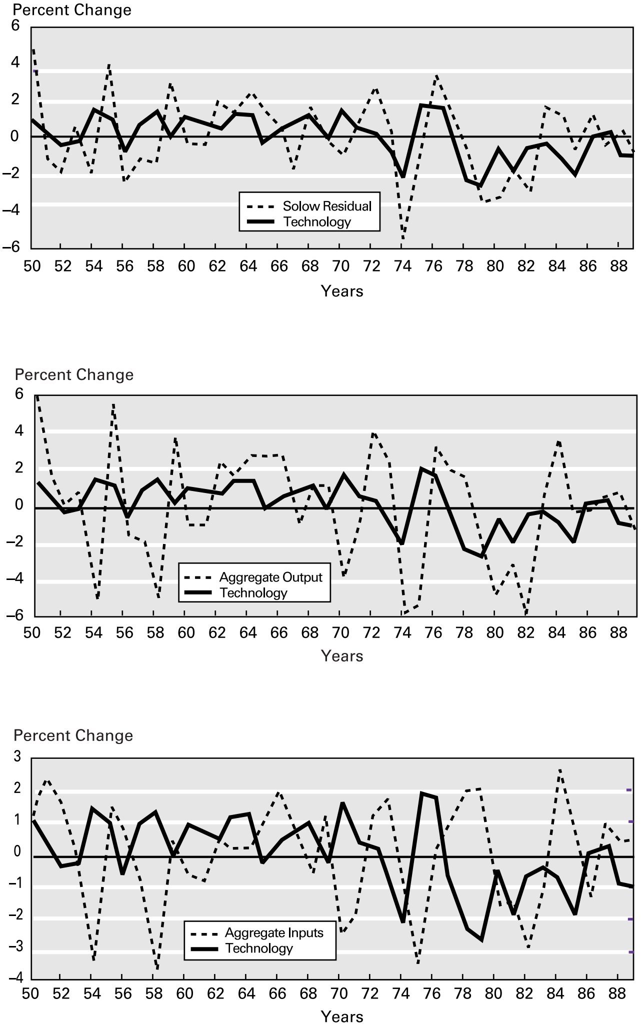
The fully corrected series now are plotted against some familiar business-cycle variables. Figures 1 and 2 show how the estimated technology series differs dramatically from the usual Solow residual.

Figure 1
Solow Residual, Input Growth, and Output Growth



Note: All series are de-meaned.

Figure 2
Technology Residual, Solow Residual, Output and Input Growth



Note: The technology series is the fully adjusted residual. All series are de-meaned.

Figure 1 plots basic business-cycle data: the Solow residual dp, aggregate output growth dv, and aggregate primary input growth, dx^V. These three series clearly co-move positively, quite strongly so in the case of dp and

dv. (All variables are de-meaned; that is, the average has been removed from all series.)

Figure 2 plots the fully corrected technology series against these three variables. Comparing technical change to the standard Solow residual, the fluctuations in the technology series are significantly smaller than the fluctuations in the Solow residual, consistent with the intuition that much of the volatility of the Solow residual reflects nontechnological factors such as variable input utilization. In addition, some periods show a phase shift: The Solow residual follows technology change with a lag of one to two years. This phase shift reflects the utilization correction: In the estimates, high technology shocks are associated with low levels of utilization, which in turn reduce the Solow residual relative to the technology series. The phase shift, in particular, appears to reflect primarily movements in hours per worker, which generally increase one year after a technology improvement. The model of the first section of this paper says that an increase in hours per worker signals an increase in unobserved effort, which the Solow residual incorrectly interprets as positive technical change.

Aggregate value-added output growth (dv) is then plotted against the same technology series. The series less clearly move together contemporaneously. Again, the series appear to have a phase shift: Output co-moves with technology, lagged one to two years. This result is qualitatively consistent with the sticky-price model in the next section, where the contemporaneous correlation between technology shocks and output growth is ambiguous but is clearly positive with a lag.

Finally, Figure 2 plots the growth rate of primary inputs of capital and labor (dx^V) and the same technology series. These two series clearly co-move negatively over the entire sample period.

It is clear that the co-movements between technology and input and output are quite different from those found in the usual real-business-cycle (RBC) literature, where one takes the standard Solow residual dp as the measure of technology change.

Why do these results differ from those of King and Rebelo (1997), who argue that variable capital utilization can "resuscitate" the RBC model? The difference arises from the different techniques used to purify the Solow residual. King and Rebelo specify a particular dynamic general-equilibrium (DGE) model, and then feed in just the observed Solow residual as data. The model then decomposes the Solow residual into technical change and variations in utilization, where the change in utilization must be consistent with the rest of the model, given the implied technology shock. In some ways this method goes into too much depth, but in other ways it is insufficiently general. For example, King and Rebelo specify a full model in order to derive the responses of labor, investment, and other variables to a technology shock. But it is not necessary to specify the environment to this extent, since these variables can all be observed in the data. On the other hand, their model does not allow for capital and labor adjustment costs, imperfect competition, sticky prices, or the variable labor effort and composition effects that we find are empirically extremely important. It is important to realize that the model of the first section nests the King-Rebelo model as a special case—the fact that I get very different results implies that the data reject their model.

Impulse Responses to Technology Improvement

I now present impulse responses of the basic variables to a technology innovation, using bivariate VARs and studying the response of a series of variables to technology shocks. The variables examined are aggregate output growth (dv), aggregate input growth (dx^V), total hours worked ($dh + dn$), and the constructed series for utilization change, du, as defined in equation (1.36).

The VAR estimated is of the form

$$\mathbf{A}(\mathbf{L})\begin{bmatrix} dt \\ dj \end{bmatrix} = \begin{bmatrix} \varepsilon \\ \eta \end{bmatrix}, \tag{3.1}$$

where dj is one of the variables studied. For dt I use the fully corrected measure of technology change. I assume that it is exogenous, so I set $a_{12}(L) = 0$. In all other cases I use a lag length of 2 periods. All equations include constants. Note that I am identifying technology shocks just as before; I am not obtaining identification from assumptions imposed on the VAR, for example, the long-run neutrality assumptions of Blanchard and Quah (1989) or Gali (1998). The VAR is just a convenient way of presenting some of the results.[19]

Figure 3 shows the impulse responses to a technology improvement: the effects of a 1 percentage point technology improvement on the (log) levels of technology, output, inputs, manhours, and utilization. Along with the impulse responses are 95 percent confidence intervals, bootstrapped using the procedure in the RATS statistical package.[20]

Both output and inputs fall on impact; the fall in inputs is strongly significant, regardless of the type of input considered (manhours, utilization, or dx^V). The fall in output is not statistically significant.

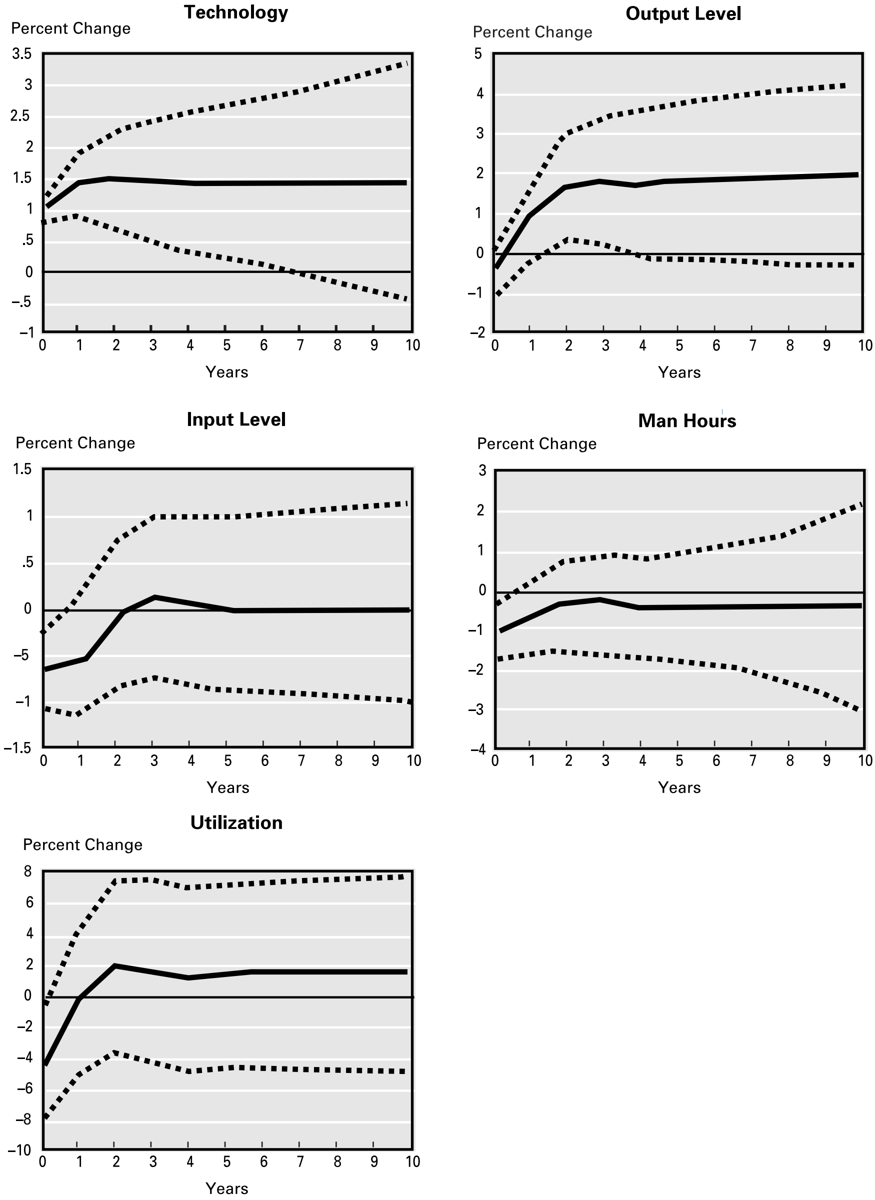
Output grows strongly after the shock; the impulse response is

---

[19] I do not use cointegration techniques, because levels of output and input need not be cointegrated with technology. For example, changes in demographic structure (for example, the baby boom) or in immigration policy can cause permanent changes in the size of the labor force that are not related to technology.

[20] These confidence intervals treat dz as data, although dz is a generated variable. They do correct for the generated-regressor problem in ε given this assumption about dz.

Figure 3
Impulse Responses to Technology Improvement: Basic Variables



Note: Impulse responses to a 1 percent age point improvement in technology. The technology series is the fully adjusted residual. Dotted lines show 95 percent confidence intervals, computed using RATS bootstrap method. Sample period is 1952 to 1989.

significantly different from zero with a two-year lag, and the point estimate shows output growing by about 1.8 percent. Inputs grow more slowly, but the standard errors of the estimates are large. For example, the point estimates say that dx^V falls 0.8 percent on impact, and then recovers to its pre-shock level (normalized to zero) in three years. However, at three years the 95 percent confidence interval runs from about 1 percent to -1 percent. The same is true of hours worked, except that the point estimate never recovers to its pre-shock value. On the other hand, the point estimates show utilization remaining above its pre-shock level indefinitely.

The finding that technology improvements reduce both output and input on impact seems problematic for standard flexible-price RBC models. This assertion will be documented below.

In a standard RBC model (for example, Cooley and Prescott 1995) with a capital share of 0.35, a 1.4 percent increase in Hicks-neutral technology (which is how I have normalized the series) should increase output by about 2.15 percent in the long run (computed as $1.4/(1 - 0.35)$), increase inputs (including capital) by about 0.75 percent in the long run, and leave manhours and utilization unchanged. The point estimate for the output response is fairly close to the predicted value. The point estimate for the input response is much lower, but the predicted value is well within the confidence interval. The same is true for utilization and hours worked.

A DGE Model with Variable Capital Utilization

This section lays out a simple sticky-price, dynamic general-equilibrium model with variable capital utilization and imperfect competition. It is representative of a number of models in the recent literature,[21] but the presentation follows Kiley (1998). The model nests a competitive flexible-price model with variable utilization and, of course, the standard RBC model with a fixed short-run supply of capital services. I do not treat variable labor effort, because in terms of model calibration the only effect of variable effort is to make the effective labor supply curve more elastic. Since I intend to follow the RBC literature and simply assume that effective labor supply is very elastic, variable labor effort is not included in the model.[22] However, variable capital utilization makes the model differ qualitatively from the standard RBC model, and this feature is discussed below. As in the empirical model of the first section of this

[21] For example, Kimball (1995); Chari, Kehoe, and McGrattan (1996); and Dotsey, King, and Wolman (1997).

[22] However, one can use the empirical evidence supporting the variable-effort hypothesis to rationalize the high short-run labor supply elasticity assumed below.

paper, the penalty for utilizing capital more intensively is that it wears out faster.

Consumers

The consumer side of the model is standard. An infinitely lived, representative consumer/worker supplies labor, rents capital, and owns the firms. The consumer maximizes the discounted value of expected utility, which is given by

$$E_t \sum_{i=0}^{\infty} \beta^i [(1-\alpha)\log(C_{t+i}) + \alpha \log(1-N_{t+i})], \tag{4.1}$$

subject to the usual series of budget constraints:

$$C_t + A_{t+1} = W_t N_t + (1+r_t)A_t + \Pi_t.$$

C is consumption, A is the consumer's stock of assets (equal to the capital stock K in equilibrium), N is labor supply, r is the real interest rate on bonds (equal to the marginal revenue product of capital minus depreciation), W is the real wage, and Π is economic profit (if any).

Optimization implies that the consumer is indifferent between consumption and leisure at a point in time, and between consumption at two different times. Thus, in equilibrium,

$$\alpha(1-N_t)^{-1} = (1-\alpha)W_t C_t^{-1}, \tag{4.2}$$

and

$$\beta E_t \left[\left(\frac{C_t}{C_{t+1}} \right) (1 + r_{t+1}) \right] = 1.$$

Since no government consumption is included in this model, aggregate output equals the sum of consumption and investment:[23]

$$Y_t = C_t + I_t. \tag{4.3}$$

The Final Goods Sector

The final goods sector is competitive and has flexible prices. The production function for final goods output, Y, uses intermediate goods,

[23] Abusing notation slightly, I use Y to denote aggregate national output. In the empirical section above, Y represented gross output, while national output is, of course, real value added (which was called V in the first section).

Y_i, as inputs to production. There is a continuum of such goods, indexed by $i \in [0, 1]$. Thus,

$$Y = \left[\int_0^1 Y_i^{\theta} \, di \right]^{1/\theta}, \qquad 0 < \theta \leq 1. \tag{4.4}$$

Note that final goods are produced with constant returns to scale. Let S_i denote the price of intermediate goods, Y_i. Then cost-minimization by the final goods firms implies constant elasticity of substitution demand functions of the form:

$$Y_i = Y \left(\frac{S_i}{P} \right)^{-1/(1-\theta)}, \tag{4.5}$$

where P is the ideal aggregate price index:

$$P = \left[\int_0^1 S_i^{\theta/(\theta-1)} \, di \right]^{(\theta-1)/\theta}.$$

Note that the monopoly markup resulting from this demand specification is constant at

$$\mu = \frac{1}{\theta}.$$

The Intermediate Goods Sector

The intermediate goods sector comprises a continuum of firms, each of which is a monopolist in the production of a single variety of good. Each firm has the production function:

$$Y_{it} = T_t (Z_{it} K_{it})^{\alpha} N_{it}^{1-\alpha} - \Phi, \tag{4.6}$$

where $\Phi \geq 0$ is a fixed cost of production (paid in units of output) and T is the level of technology. All firms have the same technology. Log technology change follows an autoregressive process:

$$\tilde{T}_t = \xi \tilde{T}_{t-1} + \omega_t. \tag{4.7}$$

If $\Phi > 0$, then firms produce with increasing returns to scale. Note, however, that increasing returns are not introduced by having diminishing marginal cost, a specification that is standard in the literature but has little empirical support. Here, marginal cost is independent of firm-level output. In this model, the steady-state degree of returns to scale is given by

$$\gamma^* = \frac{Y^* + \Phi}{Y^*}.$$

Since the increasing returns are internal to the firm, increasing returns require imperfect competition, for the reasons discussed above following equation (1.5).

The capital stock at each firm evolves according to

$$K_{i,t+1} = I_{it} + (1 - \delta(Z_{it}))K_{it}. \tag{4.8}$$

Note that unlike the empirical model, no adjustment costs for capital or employment are included. Excluding adjustment costs makes the model simpler, and also easier to compare to the existing literature.

In the sticky-price model, firms are required to set the same nominal price for two periods. Half the firms set prices in odd-numbered periods, and the other half in even-numbered periods. This specification, a variant of that introduced by Taylor (1980), is intended to capture price stickiness in a parsimonious fashion.[24] Firms set prices to maximize discounted profits. Let λ denote the firm's real marginal cost of production, as in the first section. Using a tilde (~) to denote log deviations from the steady state and assuming the one-period discount factor is approximately equal to 1, the log-linearized pricing equation is

$$\tilde{S}_{it} = \tfrac{1}{2} E_t[\tilde{P}_t + \tilde{\lambda}_{it} + \tilde{P}_{t+1} + \tilde{\lambda}_{i,t+1}]. \tag{4.9}$$

Not surprisingly, equation (4.9) shows that nominal prices are set as a markup over nominal marginal cost in the two periods. It is assumed that firms must meet all demand at the posted price; that is, rationing is ruled out. In equilibrium, all firms at time t set the same price S_t.

Note that by definition the (log) change in real marginal cost is the change in the relative price minus the change in the markup:

$$\tilde{\lambda}_{it} = (\tilde{P}_{it} - \tilde{P}_t) - \tilde{\mu}_{it}. \tag{4.10}$$

Finally, the price level is the average of prices set at t and $t - 1$:

$$\tilde{P}_t = \tfrac{1}{2}[\tilde{S}_t + \tilde{S}_{t-1}]. \tag{4.11}$$

[24] The other common variant is the Calvo/Rotemberg partial-adjustment model. Kiley (1997) compares the two specifications and argues that partial adjustment imposes a large amount of exogenous price stickiness in the case where prices are endogenously fairly flexible.

Money

Following Kiley (1998), money is introduced via an interest rate rule, where the monetary authority sets the nominal interest rate. Kiley (1998) discusses the advantages of this specification as opposed to an explicit model of money demand arising from either the presence of money in the utility function or a cash-in-advance constraint. Shocks to the interest rate rule occur, as in the empirical VAR literature. The log-linearized rule is:

$$\tilde{i}_{t+1} = \phi_y \tilde{Y}_t + \phi_p \Delta \tilde{P}_t + \nu_t, \tag{4.12}$$

where

$$\nu_t = \rho \nu_{t-1} + \varepsilon_t$$

and ε is an iid shock.

Given the nominal interest rate, the real interest rate follows from the Fisher equation:

$$\tilde{r}_{t+1} = \tilde{i}_{t+1} - E_t \Delta \tilde{P}_{t+1}.$$

Implications

Here I discuss the implications of the major innovation in the model, variable capital utilization. First, note that the cost-minimization problem facing the intermediate-goods firms in this model is a simplified version of the problem discussed in the first section. Firms face the same decision regarding variable capital utilization, although there are no investment adjustment costs and no variations in labor effort. Thus, equation (1.27) applies directly. Log-linearizing (1.27) for the case of $q \equiv 1$ and using equation (4.10), we find the expression for optimal utilization:

$$\tilde{Z}_{it} = \frac{(\tilde{P}_{it} - \tilde{\mu}_{it}) + (1/\gamma^*)\tilde{Y}_{it} - \tilde{K}_{it}}{1 + \Delta}. \tag{4.13}$$

As before, Δ is the elasticity of the *marginal* rate of depreciation with respect to utilization.

Substituting equation (4.13) into the log-linearized production function, we find:

$$\tilde{Y}_{it}\left(1 - \frac{\alpha}{1 + \Delta}\right) = \frac{\gamma^* \alpha \Delta}{1 + \Delta} \tilde{K}_{i,t-1} + \gamma^*(1 - \alpha)\tilde{N}_{it} + \frac{\gamma^* \alpha}{1 + \Delta}((\tilde{P}_{it} - \tilde{P}_t) - \tilde{\mu}_{it}) + \gamma^* \tilde{T}_t. \tag{4.14}$$

As King and Rebelo (1997) observe, one gains intuition about the effects of variable capital utilization by studying the limiting cases of equation

(4.14). First, suppose that $\Delta = \infty$. Then changes in capital utilization are so costly that utilization never changes, and (4.14) reduces to the familiar equation for log-linearized output growth with increasing returns:

$$\tilde{Y}_{it} = \gamma^* \alpha \tilde{K}_{i,t-1} + \gamma^*(1-\alpha)\tilde{N}_{it} + \gamma^* \tilde{T}_t.$$

Second, suppose that $\Delta = 0$, so that depreciation increases only linearly with utilization. Suppose we are in the perfectly competitive case, where $(\tilde{P}_{it} - \tilde{P}_t) - \tilde{\mu}_{it} \equiv 0$. Then the reduced-form production function is linear in labor input—thus, there is effectively no diminishing marginal product of labor, even in the short run, implying that marginal cost is less procyclical and the propagation mechanism for external shocks is stronger. Finally, in the imperfectly competitive case, note that a firm with countercyclical markups experiences larger changes in output if Δ is small. For example, if monetary policy is unexpectedly expansionary and all prices are sticky, the change in the relative price is zero but the change in the markup is negative.[25] This countercyclical markup has the expansionary effect of increasing the demand for all factors, including capital services. The resulting increase in utilization is larger if Δ is small.

Calibration

This subsection discusses the calibration of the RBC model with variable capital utilization, and the additional parameters needed to calibrate the sticky-price model. The calibration is done so that each model period corresponds to two quarters. Thus, in the sticky-price model, each firm keeps its price fixed for one year.

The RBC Model. The RBC model consists of the model above with one-period price setting (thus making prices perfectly flexible), perfect competition, and constant returns. Perfect competition requires $\theta = 1$; constant returns implies that $\Phi = 0$. With two exceptions, the remaining parameters are calibrated to equal those of the benchmark RBC model of Cooley and Prescott (1995, p. 22).[26] In particular, the critical intertemporal

[25] The markup falls because expansionary monetary policy necessarily raises marginal cost in this model. In other models, marginal cost might actually fall as output increases—for example, in the "sunspot" model of Farmer and Guo (1994). Markups might also be countercyclical in flexible-price models, for reasons advanced by Rotemberg and Woodford (1992) and Gali (1994).

[26] Since this model does not have steady-state growth, I set the investment share to match the data using equation (33) in Cooley and Prescott (1995). Thus, although the model has neither trend technology growth nor population growth, I use the values for those parameters found in Cooley and Prescott's table (1995, p. 22). The model would be quite consistent with steady-state growth with some modification that makes the degree of returns to scale stationary. One such change would be to have the size of fixed costs grow deterministically at the trend rate of growth of the economy, but there are also other possibilities. See Rotemberg and Woodford (1991, 1995).

elasticity of labor supply is calibrated to equal approximately 2.2 (somewhat lower than most of the RBC literature, though still higher than most of the micro estimates of this parameter).

The first exception is the variance of the innovation to technology. As noted above, I use the actual innovations to technology estimated using the BFK procedure.[27] However, since the estimated residuals are annual, I assume that the technology shock occurs in the first half of each year. The technology shock for the second half of the year is always zero. Since agents in the model always expect the technology innovation to equal zero, and the log-linearization eliminates higher-order responses to uncertainty such as precautionary saving, this procedure does not cause any obvious problems.

The second exception is the parameter Δ, which Cooley and Prescott (1995) do not need to calibrate since their model implicitly assumes $\Delta = \infty$. Burnside and Eichenbaum (1996) calibrate $\Delta = 0.56$, but they do so using a very restrictive functional form that implies

$$\Delta \equiv \frac{Z^*\delta''(Z^*)}{\delta'(Z^*)} = \frac{Z^*\delta'(Z^*)}{\delta(Z^*)} - 1 = \frac{r^*}{r^* + \delta(Z^*)}. \tag{4.15}$$

This method thus identifies Δ purely from a functional form assumption, which is clearly undesirable. Basu and Kimball (1997) discuss the shortcomings of this approach. They estimate Δ from an instrumental variables regression of equation (1.33) and find Δ approximately equal to 1, but with a large standard error (also about 1). I therefore use 1 as my benchmark value of Δ, but also experiment with other values.

Finally, the log-linearized capital accumulation equation also requires one to calibrate the elasticity of $\delta(Z^*)$. However, as equation (4.15) shows, this parameter can be calibrated from the steady-state real rate of interest (which in turn is a function of the discount rate β) and the steady-state depreciation rate, and does not require additional estimation.

The Sticky-Price Model. The real side of the sticky-price model is identical to that of the RBC model, with one exception: the degree of imperfect competition and markups. I follow a calibration that assumes zero economic profit in the steady state and thus equates the markup and the steady-state degree of returns to scale (see equation 1.5).[28] A variety of evidence indicates that the plausible degree of imperfect competition

[27] However, I do maintain the Cooley-Prescott calibration of 0.95 (quarterly) for the autoregressive parameter ξ. The point estimate for ξ is actually larger than 1, but one cannot reject the lower value at the 95 percent level. I thus maintain the Cooley-Prescott value for the model simulations, to allow easier comparison to the existing literature.

[28] Rotemberg and Woodford (1995) present a variety of evidence supporting the proposition that pure profit rates are close to zero.

and/or the degree of increasing returns is small.[29] Thus, I set $\gamma^* = \mu^* = 1.05$, implying that $\theta = 0.95$ and $\Phi/Y^* = 0.05$. Recall that this markup, already very small, is the markup on real value added. Assuming a material's share in production of 0.50, it implies that firms sell actual goods for about 2 percent higher than their marginal cost of production—a calibration well within the confidence interval of any recent estimate.

The sticky-price model also requires another set of parameters, relating to the nominal side of the model. Interpreting equation (4.12) as a policy rule, Taylor (1993) suggests that in the post-1987 period the Federal Reserve has followed a policy described by setting $\phi_y = 0.5$ and $\phi_p = 1.5$. These are the parameters I adopt as my baseline case, as does Kiley (1998). Since it is unlikely that the Fed has adhered to this rule over the 40 years of my sample period, the historical simulations based on this rule should be treated as suggestive. I also experiment with reinterpreting equation (4.12) as a standard LM curve, with an exogenous money supply, an income elasticity of money demand equal to 1, and an interest elasticity of money demand equal to -0.5.[30] The calibration implies $\phi_y = 2$ and $\phi_p = 2$. The steady-state inflation rate is assumed to be zero.

In all cases, the autoregressive parameter ρ is set to 0.50.[31] The impulses for the money supply rule are residuals from an assumed monetary policy reaction function, estimated by Burnside (1996).[32] They are residuals from an OLS regression of the 3-month T-bill rate on lags of itself and on current and lagged values of GDP growth, inflation, and commodity prices. The estimation is done at a quarterly frequency, so the shocks for the first period are the sum of the shocks in the first two quarters and the shocks for the second period are the sum of the shocks from the third and fourth quarters.

Simulation Results

This section presents two sets of results: impulse responses to both technology and monetary policy shocks, and historical simulations, of the sort performed by Plosser (1989).

Impulse Responses

I first present results for the RBC model, the benchmark model described above, with variable capital utilization. The value of Δ is set to

---

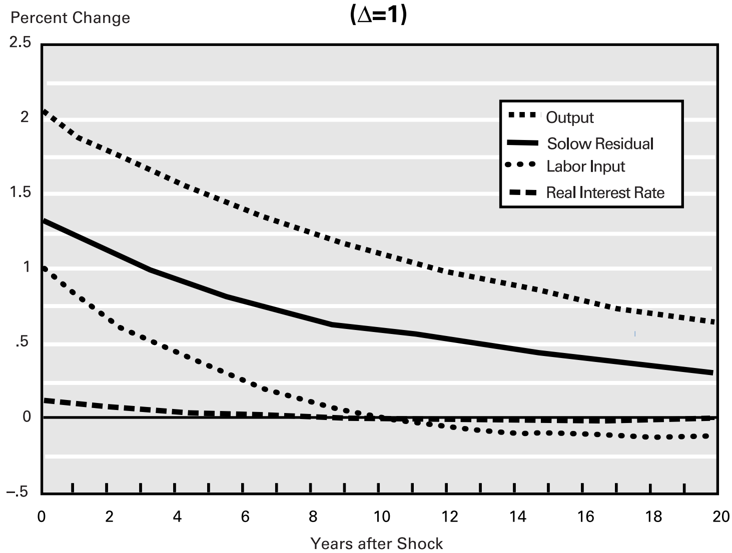
[29] See, for example, Basu (1996), Burnside (1996), and Basu and Fernald (1997).

[30] In this case the equation gives the value of the nominal interest rate at time t, not $t + 1$.

[31] See Kiley (1998) for a discussion.

[32] I thank Craig Burnside for providing these data.

Figure 4
Impulse Response to a 1-Percentage-Point Technology Improvement in the RBC Model
(Δ=1)



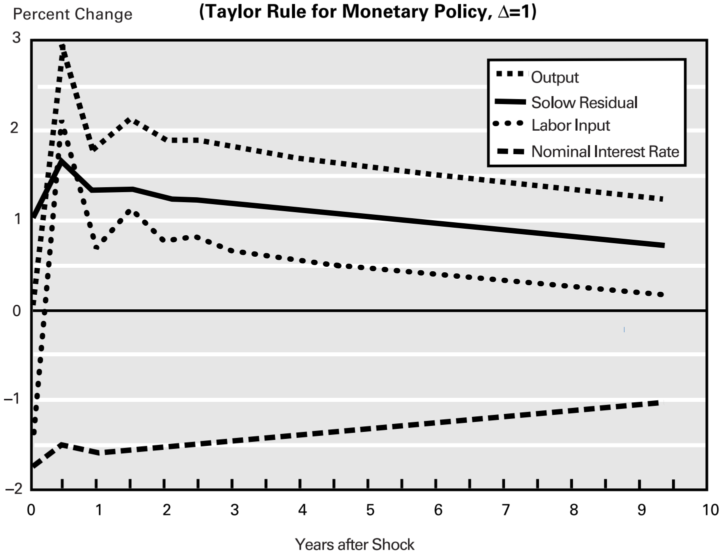
the estimate of 1 in Basu and Kimball (1997). The series shown are the responses of output, labor hours, the real interest rate, and the Solow residual. The residual is calculated as it would be from the data—that is, changes in utilization show up as changes in the residual. Knowing the time path of technology, we can infer the time series for utilization.

Figure 4 gives the impulse response to a 1-percentage-point technology improvement in the RBC model. Output rises by about 2 percent on impact, and labor input by about 1 percent. Note that the Solow residual is higher than 1 on impact, showing that capital utilization increases in response to the technology improvement. Thus, as conjectured, variable utilization amplifies the effects of shocks.

But, as we know from a long line of work, the impulse responses of the RBC model are dramatically different from the empirical results presented in Figure 3. In the data, both output and labor input fall when technology improves, and reach their peak two or three years later. The model shows no fall on impact; all variables are at their peak at time zero.

We now turn to the sticky-price model, to see whether it can explain the observed impulse responses. Figure 5 presents the sticky-price version of the RBC model simulated above. The model has the bench-

Figure 5
Impulse Response to a 1-Percentage-Point Technology Improvement in the Sticky-Price Model
(Taylor Rule for Monetary Policy, Δ=1)



mark calibration, including $\Delta = 1$ and monetary policy as described by the Taylor rule. (The figures for the sticky-price model display the time path for the nominal interest rate, rather than the real interest rate.)

The results are mildly encouraging. Labor hours fall significantly on impact—about 1.5 percent, more than the point estimate from the data but well within the confidence interval. Output basically does not respond on impact, which is also quite consistent with the data. The Solow residual rises by less than 1 percent, showing that utilization must have fallen, as it does in the data. The model thus displays the strong negative co-movement between technology and inputs that is observed in Figure 3.

The major failure of the sticky-price model is its inability to explain the drawn-out contraction observed in the data. Output and inputs reach their peak just one period (six months) after the shock, when only half the firms have changed prices. They then fall somewhat, before converging smoothly to the steady state. The empirical results, however, have output and inputs reaching their peak about two or three years after the shock.[33]

[33] Part of this pattern may be changed by using the actual ARI(1,1) process for technology change that is observed in the data.

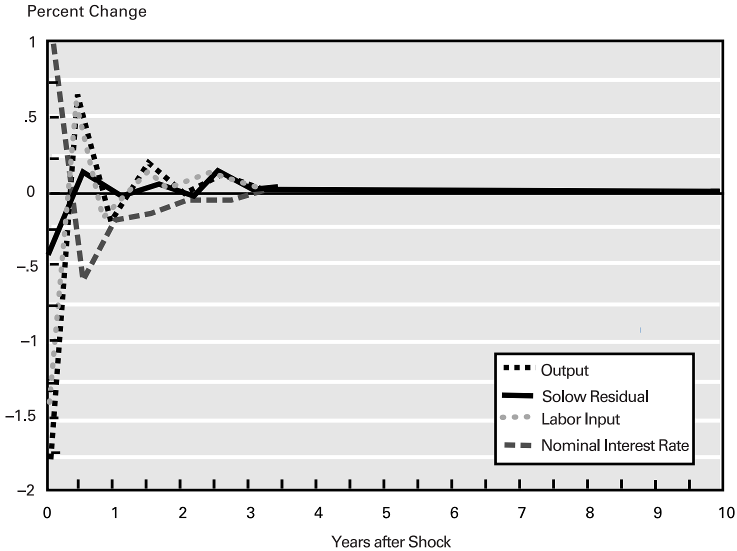
I now turn to the other major issue to be investigated, the importance of variable utilization as a propagation mechanism. I first study the effects of variable utilization on the impulse response for nominal shocks, and return to technology shocks after this detour. Figure 6 shows the effects of a 1-percentage-point increase in the nominal interest rate in the model just simulated. The experiment is to shock ν_t in equation (4.12) by 1 percentage point, and then let the future path of the nominal interest rate be given by the autoregressive time path for ν, as well as the endogenous monetary response dictated by the Taylor rule. The results in Figure 6 are not encouraging. Output falls by more than 1.5 percent in the period of the shock, as does labor input. The Solow residual falls as well, matching the co-movement between observed productivity and nominal shocks documented by Evans (1992). In this case, the majority of the fall is due to the reduction in utilization, though a small percentage can be attributed to the effects of increasing returns to scale in production. With increasing returns, productivity changes when inputs change, and in the same direction. Note that the Taylor rule dictates very expansionary monetary policy in response to the fall in output and inflation. Even though the exogenous component of monetary policy is still tight, the endogenous response is so large that the nominal interest rate falls to -0.56 percent.

However, the shock is not propagated over time. Indeed, output and inputs "overshoot" the steady state only one period after the shock, and converge quickly to the steady state in an oscillatory fashion. This result is puzzling given the encouraging findings of Dotsey, King, and Wolman (1997), who calibrate a similar model, with the same parameters governing variable utilization, and report moderate persistence eight quarters after a nominal shock. However, they assume that the monetary authority does not respond to an economic contraction by loosening monetary policy: The nominal interest rate in their model is derived from an LM specification for money demand. The behavior of the nominal interest rate in Figure 6 suggests that endogenous monetary policy is quite important. I thus change the calibration of this part of the model, in an effort to see how much of the difference in my results comes from the assumption of endogenous monetary policy. I replace the Taylor rule with an LM curve, assuming that the elasticity of money demand with respect to output is 1, and its elasticity with respect to the nominal interest rate is -0.5. (The latter figure is probably somewhat high given the empirical estimates in the literature.)

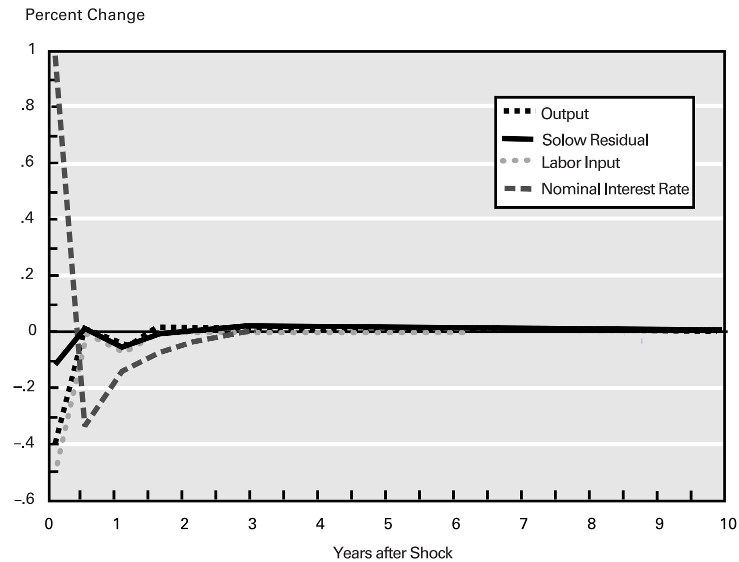
Figure 7 reports the results. The experiment is still a 1 percentage point increase in ν, but now the nominal interest rate falls only -0.3 percentage points below its steady-state level the year after a shock. The effects of the increase in the interest rate are much smaller—output and inputs fall less than 0.5 percentage points—but this calibration avoids the overshooting result: The period after the shock, output and inputs are

Figures 6 and 7
Impulse Response to a Monetary Contraction[a]
in the Sticky-Price Model

Taylor Rule for Monetary Policy, Δ=1



Exogenous Money, Δ=1



[a] 1-percentage-point increase in nominal interest rate.

back at the steady state, not significantly higher than their steady-state levels. I still do not reproduce the Dotsey et al. results, but I am using much higher values for the interest elasticity of money demand (they use zero) and the labor supply elasticity (which they assume is infinite).

However, as Kiley (1998) argues, it seems better to calibrate sticky-price models using Fed reaction functions rather than money demand equations that assume exogenous monetary policy.[34] First, an interest rate targeting function is clearly a more realistic description of how Fed policy now operates; the Fed definitely perceives the nominal interest rate as its policy instrument. Second, the results in Dotsey et al. and Chari, Kehoe, and McGrattan (1996) are sensitive to the assumed interest elasticity, which is a very poorly estimated parameter. The advantage of the reaction-function approach is that the money demand parameters are irrelevant for the results. Thus, the challenge is to reproduce the Dotsey et al. results, using the more realistic framework employed here.

One method is to follow Dotsey et al. and make labor supply infinitely elastic. This strategy seems problematic—the elasticity of 2 assumed here seems about as high as one can reasonably get, even with variable labor effort. The other way to make factor supply more elastic is to reduce the size of Δ. Since Δ is quite imprecisely estimated, it seems reasonable to experiment with values smaller than one.

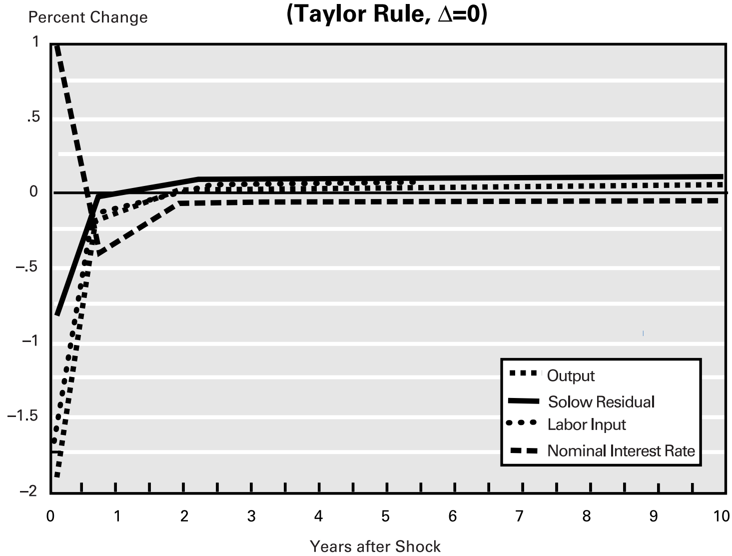
Figure 8 reports the results for the limiting case of $\Delta = 0$, returning to the case where the nominal interest rate is set by the Taylor rule. The results seem encouraging; output and inputs fall by almost 2 percent and remain above their steady-state levels for a year. No overshooting occurs, as was the case for the same model with $\Delta = 1$ (shown in Figure 6).

However, there is still not enough persistence, relative to the results found in the empirical literature. Of course, this result may simply indicate still not enough "real rigidities" in the model. Ball and Romer (1990) point out the importance of real rigidities for generating substantial real effects of nominal shocks in static models with state-dependent pricing; Kimball (1995) confirms their results in a dynamic DGE model with time-dependent pricing. In the context of this model, the degree of real rigidity corresponds (inversely) to the assumed size of Δ, given the other calibration, particularly the intertemporal elasticity of labor supply. Thus, the answer may simply be that we need even more real rigidity than the minimum value of Δ allows. For example, countercyclical variation in the size of *desired* markups would help enormously, for reasons explained by Kimball (1995).

However, the fact that we now want sticky-price models to generate

[34] However, it is not clear that the Taylor rule is the best one. Orphanides (1997) argues that estimation using real-time data supports simple forward-looking rules over the Taylor rule.

Figure 8
Impulse Response to a Monetary Contraction[a]
in the Sticky-Price Model
(Taylor Rule, Δ=0)

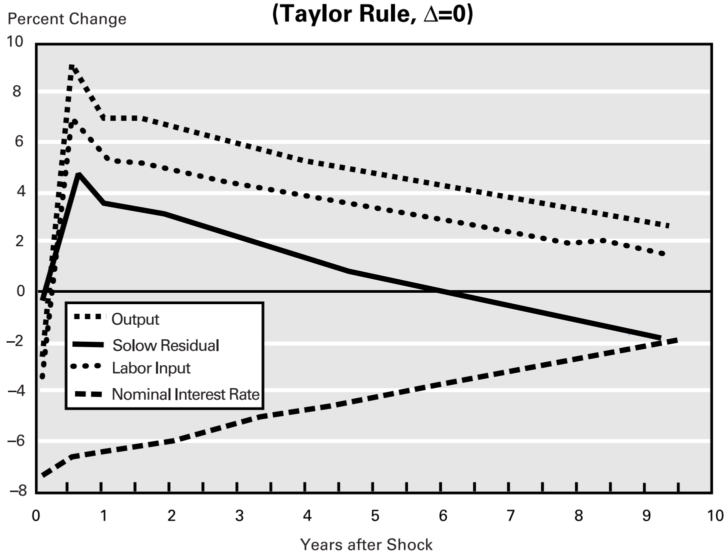


[a] 1-percentage-point increase in nominal interest rate.

sensible impulse responses to technology shocks puts greater constraints on the search for real rigidities. To document this assertion, Figure 9 shows what happens in the sticky-price model with $\Delta = 0$ in response to a 1-percentage-point technology improvement. First, we see the expected fall in inputs and output, now almost 4 percent. Utilization falls so much that the Solow residual actually *falls* slightly in response to an *improvement* in technology! But then, in the first period after the shock, the model predicts enormous increases in output (almost 10 percent) and inputs (about 7 percent). These far exceed in magnitude any of the impulse responses observed in the data, at any lag. But, according to the fully adjusted technology residual, a 1-percentage-point or larger change in technology relative to its mean is not an uncommon event in the data—we find such changes in 17 of the 40 years of the sample. Thus, the $\Delta = 0$ model must be rejected for implying *too much* real rigidity to be consistent with the observed effects of technology—while at the same time it generates *too little* real rigidity to rationalize persistent real effects of money.

The reasons are not hard to understand. A nominal shock is fundamentally a weak shock, relying on price rigidity to have any real

Figure 9
Impulse Response to a 1-Percentage-Point Technology Improvement in the Sticky-Price Model (Taylor Rule, Δ=0)



effects. For weak shocks to have large, persistent effects, the economy needs to be in a state of "near indeterminacy," depicted in the labor market as flat labor supply and demand curves lying almost on top of one another.[35] But technology shocks are large, real shocks. An economy that displays business-cycle-sized fluctuations in response to money shocks may well display implausibly large fluctuations in response to technology shocks—as Figures 8 and 9 show.

This result does not mean that the search for a sensible, integrated model of business cycles is hopeless. What it does imply is that we need "shock-dependent" real rigidities. For example, the implicit-collusion model of countercyclical markups[36] predicts that changes in the markup depend on the time path of output and interest rate responses to shocks. These paths are likely to differ in response to different shocks. Similarly, if leisure is durable, then labor supply will be more elastic in response to

[35] The phrase and interpretation are from Hall (1991). Kimball (1995) and Kiley (1998) discuss this issue in the context of sticky-price DGE models.

[36] Rotemberg and Woodford (1992, 1995).

temporary rather than to permanent changes in the demand for labor. If the effects of monetary shocks last only over the "short run," but the effects of technology shocks last over the "medium run," then the degree of real rigidity from flat labor supply may well be larger in the case of monetary shocks.

Historical Simulations

I now simulate the two benchmark models using historical shocks, and then compare the realizations with actual data for the years 1950 to 1989. The simulation for the RBC model is simple, since I need only choose a value for Δ (set equal to one) and feed in the estimated series of technology shocks. The sticky-price model simulation is much harder, since I also need to specify the form of the Fed's reaction function. For a historically accurate simulation, one would actually need to supply a variety of reaction functions, since it seems clear that the Fed has fundamentally changed its procedures several times over this sample period. However, I assume that the reaction function was given by the Taylor rule throughout, but the exogenous component of money was subject to the shocks estimated by Burnside's (1996) VAR. Thus, the simulation should be regarded more as an instructive exercise than a rigorous attempt to duplicate the historical record. I also maintain $\Delta = 1$ for this model.

The results for the RBC model are summarized in Table 2. The first panel shows basic standard deviations and correlations for the data. For the data, I use private output growth (a chain-weighted index of aggregate GDP minus government purchases); private consumption, investment, and hours worked; and the Solow residual for the private economy. The only surprises are the high correlation of consumption growth with output growth, which I find to be 0.92, and the high standard deviation of consumption, 1.81. Most studies using annual data put these figures at about 0.8 and 1.3. I speculate that the difference comes partly from the fact that I am using the new chain-weighted NIPA data and partly from my definition of output, which excludes government. On the other hand, I find that investment is somewhat less correlated with output than generally reported.

As one might have guessed from the impulse responses, the results are not kind to the RBC model. Since the volatility of the estimated technology shocks is much smaller than the volatility of the Solow residual series used by Plosser (1989), the model underpredicts the standard deviations of all the variables. All variables are too highly correlated with output, a standard result when only one shock is driving all fluctuations. Most problematically, the correlations of the simulated growth rates with the actual ones are mostly negative. The correlation between the actual and simulated output series is -0.21, and the

Table 2
Summary Statistics for Historical Simulation: RBC Model
($\Delta = 1$)

A. Actual

	Standard Deviation	Correlation with Output	Correlation with Actual
$\Delta \log (Y)$	2.97	1.00	1.00
$\Delta \log (C)$	1.81	.92	1.00
$\Delta \log (I)$	10.51	.86	1.00
$\Delta \log (N)$	2.28	.80	1.00
$\Delta \log (SR)$	2.98	.98	1.00

B. Predicted

	Standard Deviation	Correlation with Output	Correlation with Actual
$\Delta \log (Y)$	2.30	1.00	−.21
$\Delta \log (C)$	.98	.94	.08
$\Delta \log (I)$	6.44	.99	−.41
$\Delta \log (N)$	1.16	.97	−.62
$\Delta \log (SR)$	1.48	1.00	−.12

correlation between the two labor series is −0.62. Time-series plots for the actual and generated series (in de-meaned growth rates) are shown in Figure 10.

Results for the sticky-price model are more encouraging (Table 3). Many of the standard deviations are much higher; in fact, the standard deviation of hours worked is almost 50 percent larger in the model than in the data. This result is not typical of DGE models, particularly ones with such low labor supply elasticities as the model used here. Hours are so volatile for two reasons. First, capital utilization is allowed to vary, reducing the rate at which the diminishing marginal product of labor sets in. Second, technology shocks produce a "whiplash" effect, first reducing then increasing hours above their steady-state level. The correlations with output are also reduced—implausibly so, in the case of consumption. (The main reason seems to be that consumption rises in response to a technology improvement, even though output falls on impact.) Most importantly, the correlations with the actual series are positive, albeit mildly so. For example, the output and hours correlations are both about 0.3.

This model seems promising, because two small modifications are likely to go a long way towards improving the summary statistics. The first is adding variable labor effort. I have argued that from a modeling standpoint variable effort is equivalent to a higher labor supply elasticity. Thus the labor series in the model should be regarded as the sum of observed (hours) and unobserved (effort) labor fluctuations in the model.

Figure 10
Historical Simulation: RBC Model
(Δ=1)

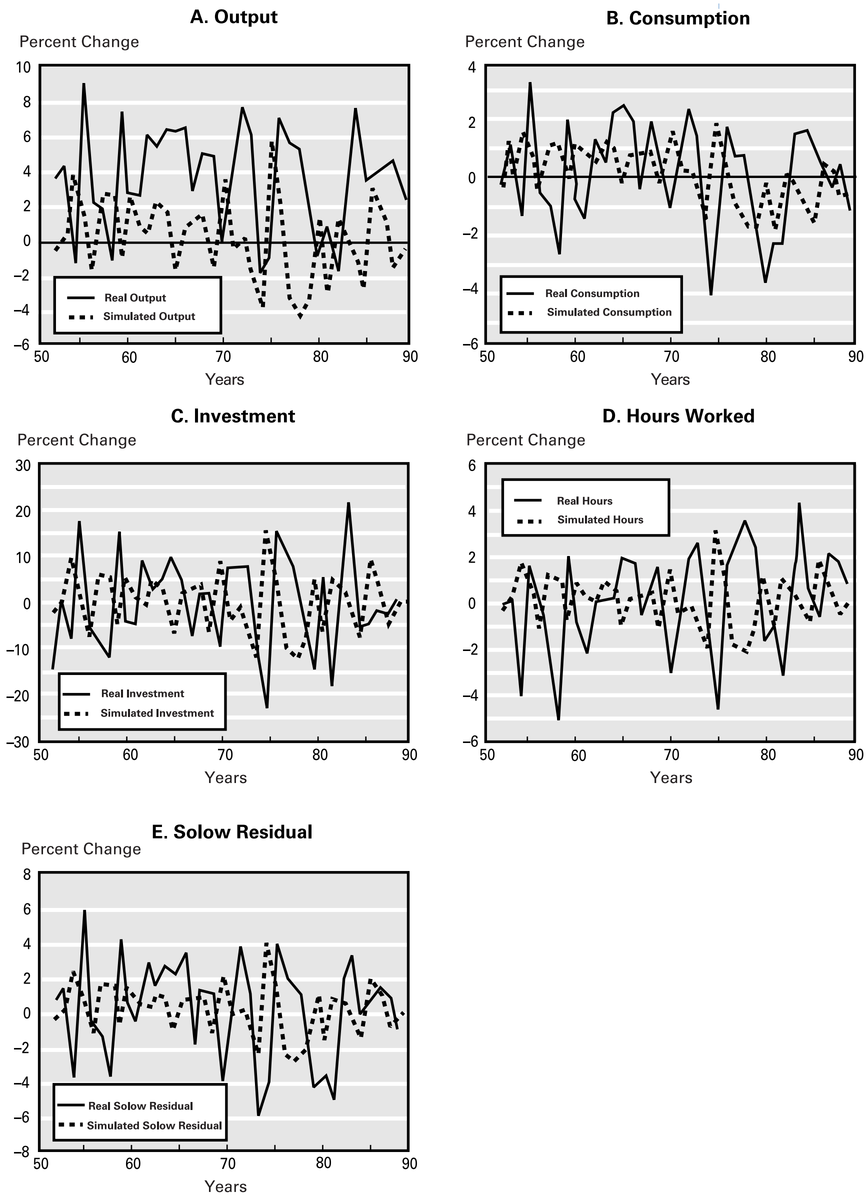


Table 3
Summary Statistics for Historical Simulation: Sticky-Price Model
(Taylor Rule, $\Delta = 1$)

A. Actual

	Standard Deviation	Correlation with Output	Correlation with Actual
Δ log *(Y)*	2.97	1.00	1.00
Δ log *(C)*	1.81	.92	1.00
Δ log *(I)*	10.51	.86	1.00
Δ log *(N)*	2.28	.80	1.00
Δ log *(SR)*	2.98	.98	1.00

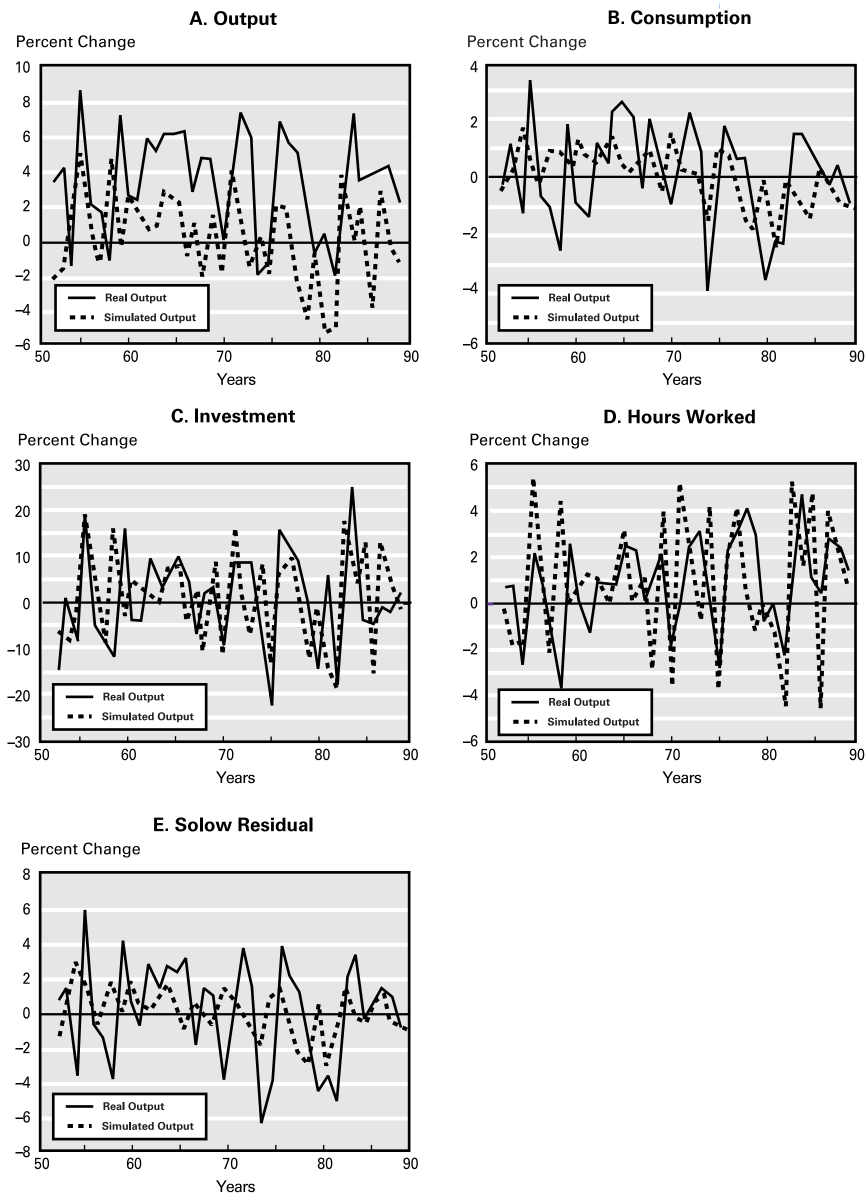
B. Predicted

	Standard Deviation	Correlation with Output	Correlation with Actual
Δ log *(Y)*	2.59	1.00	.30
Δ log *(C)*	1.03	.35	.19
Δ log *(I)*	9.62	.95	.35
Δ log *(N)*	3.19	.84	.33
Δ log *(SR)*	1.29	.59	.12

But from an empirical standpoint, modeling unobserved effort would increase the standard deviation of the simulated Solow residual and its correlation with output, while reducing the standard deviation of hours—all desirable outcomes. Second, both the volatility of consumption and its correlation with output can probably be increased by modeling liquidity constraints. After all, most research on consumption strongly rejects the simple permanent-income model used in the DGE literature. Liquidity constraints might prevent the countercyclical behavior of consumption in response to technology improvements that is found in the model but not the data (see BFK).

However, the model is nowhere close to experiencing the sort of success that Plosser (1989) claimed for the RBC model. Inspecting the time-series plots in Figure 11 suggests that the reason is the lack of propagation discussed earlier. Panel A, which plots the two output series, shows that fluctuations do not seem as long-lived in the model as in the data, leading to the relatively poor fit between prediction and outcome. It may be possible to reduce this problem by adding more real rigidities to the model—but bearing in mind the caveat discussed above.

Figure 11
Historical Simulation: Sticky-Price Model
(Taylor Rule, Δ=1)



CONCLUSION

The behavior of the economy in response to technology shocks is a challenge for business-cycle theory. Current empirical results show that when technology improves, inputs fall significantly in the short run and output is almost unchanged. The results in this paper show that benchmark models of the business cycle are unable to rationalize this behavior fully. Sticky-price models show some promise of being able to match the data, but they clearly have a long way to go.

Disappointingly, just adding a plausible amount of variable factor utilization to the sticky-price model does not impart enough real rigidity to match estimated impulse responses for either technology or money. In some ways, this result may not seem particularly discouraging. After all, variable utilization is only one of many possible real rigidities, and the others may pick up the slack. Variable capital utilization is different, however, in that it is solidly documented (for example, by Shapiro 1996, using firm-level data), and some of the parameters governing changes in utilization have been estimated. By contrast, many of the other mechanisms discussed in the literature—for example, kinked demand curves, sector-specific externalities, efficiency wages, or countercyclical target markups—remain more in the realm of wishful thinking.

Finally, while adding technology shocks to sticky-price models holds the promise of being able to explain the puzzling facts about the effects of technology on the economy, researchers now face the challenge of producing sensible impulse responses for two kinds of shocks using the same model. While this discipline is desirable, it makes an already difficult job even harder.

However, a second class of propagation mechanisms that has some solid empirical support has not been considered here. These are models where, as a result of frictions of some kind, cyclical changes in the composition of output serve to magnify the effects of shocks. And since different types of shocks lead to different output composition, this class of models has the potential to produce different degrees of real rigidities for technology shocks than for money shocks. However, research on calibrated multisector models with frictions is extremely demanding, both computationally and in terms of the effort needed to understand the workings of the model at a deep level. Some research is under way,[37] but it is too early to say whether these models will make a significant contribution to solving the problems identified here.

[37] See, for example, Horvath (1995), Basu, Fernald, and Horvath (1996), and Phelan and Trejos (1996).

References

Ball, Laurence and David Romer. 1990. "Real Rigidities and the Non-neutrality of Money." *Review of Economic Studies*, *57* (April), pp. 183–203.

Basu, Susanto. 1996. "Cyclical Productivity: Increasing Returns or Cyclical Utilization?" *Quarterly Journal of Economics*, *111* (August), pp. 719–51.

Basu, Susanto and John G. Fernald. 1997. "Aggregate Productivity and Aggregate Technology." International Finance Discussion Paper 593, Board of Governors of the Federal Reserve System.

______. 1998. "Why is Productivity Procyclical? Why Do We Care?" Forthcoming in NBER volume, *New Directions in Productivity Research*.

Basu, Susanto, John G. Fernald, and Michael T. K. Horvath. 1996. "Aggregate Production Function Failures." Manuscript.

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 1998. "Are Technology Improvements Contractionary?" Manuscript.

Basu, Susanto and Miles S. Kimball. 1997. "Cyclical Productivity with Unobserved Input Variation." NBER Working Paper No. 5915.

Bils, Mark and Jang-Ok Cho. 1994. "Cyclical Factor Utilization." *Journal of Monetary Economics*, *33* (April), pp. 319–54.

Blanchard, Olivier and Daniel Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *The American Economic Review*, *79*, no. 4, pp. 654–73.

Burnside, Craig. 1996. "What Do Production Function Regressions Tell Us about Increasing Returns to Scale and Externalities?" *Journal of Monetary Economics*, *37* (April), pp. 177–201.

Burnside, Craig and Martin Eichenbaum. 1996. "Factor-Hoarding and the Propagation of Business-Cycle Shocks." *The American Economic Review*, *86*, pp. 1154–74.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 1996. "Sectoral Solow Residuals." *European Economic Review*, *40*, pp. 861–69.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan. 1996. "Sticky-Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?" NBER Working Paper No. 5809.

Cooley, Thomas F. and Edward C. Prescott. 1995. "Economic Growth and Business Cycles." In Thomas F. Cooley, ed., *Frontiers of Business Cycle Research*. Princeton, NJ: Princeton University Press.

Domar, Evsey D. 1961. "On the Measurement of Technical Change." *Economic Journal*, *71* (December), pp. 710–29.

Dotsey, Michael, Robert G. King, and Alexander Wolman. 1997. "Menu Costs, Staggered Price Setting, and Elastic Factor Supply." Manuscript, June.

Evans, Charles L. 1992. "Productivity Shocks and Real Business Cycles." *Journal of Monetary Economics*, *29* (April), pp. 191–208.

Farmer, Robert and Guo, Jang-Ting. 1994. "Real Business Cycles and the Animal Spirits Hypothesis." *Journal of Economic Theory*, *63*, pp. 42–72.

Gali, Jordi. 1994. "Monopolistic Competition, Business Cycles, and the Composition of Aggregate Demand." *Journal of Economic Theory*, *63*, pp. 73–96.

______. 1998. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *The American Economic Review*, forthcoming.

Griliches, Zvi and Jacques Mairesse. 1995. "Production Functions: The Search for Identification." NBER Working Paper No. 5067.

Haavelmo, Trygve. 1960. *A Study in the Theory of Investment*. Chicago: Chicago University Press.

Hall, Alastair R., Glenn D. Rudebusch, and David W. Wilcox. 1996. "Judging Instrument Relevance in Instrumental Variables Estimation." *International Economic Review*, *37* (May), pp. 283–98.

Hall, Robert E. 1988. "The Relation Between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy*, *96* (Oct.), pp. 921–47.

______. 1990. "Invariance Properties of Solow's Productivity Residual." In Peter Diamond, ed., *Growth, Productivity, Unemployment*. Cambridge, MA: The MIT Press.

______. 1991. "Labor Demand, Labor Supply, and Employment Volatility." In Olivier Blanchard and Stanley Fischer, eds., *NBER Macroeconomics Annual*.

Hall, Robert E. and Dale W. Jorgenson. 1967. "Tax Policy and Investment Behavior." *The American Economic Review*, *57* (June), pp. 391–414.

Horvath, Michael T. K. 1995. "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks." Manuscript, Stanford University.

Hulten, Charles. 1978. "Growth Accounting with Intermediate Inputs." *Review of Economic Studies*, *45*, pp. 511–18.

Jorgenson, Dale W., Frank Gollop, and Barbara Fraumeni. 1987. *Productivity and U.S. Economic Growth*. Cambridge, MA: Harvard University Press.

Jorgenson, Dale W. and Zvi Griliches. 1967. "The Explanation of Productivity Change." *Review of Economic Studies*, *34*, pp. 249–83.

Jorgenson, Dale W. and Kun-Young Yun. 1991. *Tax Reform and the Cost of Capital*. Oxford: Oxford University Press.

Kiley, Michael T. 1997. "Partial Adjustment and Staggered Price Setting." Manuscript, Federal Reserve Board.

______. 1998. "Staggered Price Setting and Real Rigidities." Manuscript, Federal Reserve Board, February.

Kimball, Miles S. 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit, and Banking*, *27* (November), pp. 1241–77.

King, Robert G. and Sergio Rebelo. 1997. "Resuscitating Real Business Cycles." Manuscript, December.

Lilien, David M. 1982. "Sectoral Shifts and Cyclical Unemployment." *Journal of Political Economy*, *90* (August), pp. 777–93.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, *64*, pp. 1263–97.

Orphanides, Athanasios. 1997. "Monetary Policy Rules Based on Real-Time Data." Manuscript, Federal Reserve Board, December.

Phelan, Christopher and Alberto Trejos. 1996. "On the Aggregate Effects of Sectoral Reallocation." Manuscript, Northwestern University.

Plosser, Charles I. 1989. "Understanding Real Business Cycles." *Journal of Economic Perspectives*, *3* (Summer), pp. 51–77.

Prescott, Edward C. 1986. "Theory Ahead of Business Cycle Measurement." *Carnegie Rochester Conference Series on Public Policy*, *25* (Autumn), pp. 11–44.

Rotemberg, Julio J. and Michael Woodford. 1991. "Markups and the Business Cycle." *NBER Macroeconomics Annual*.

______. 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy*, *100* (December), pp. 1153–1207.

______. 1995. "Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets." In Thomas F. Cooley, ed., *Frontiers of Business Cycle Research*. Princeton, NJ: Princeton University Press.

Schor, Juliet B. 1987. "Does Work Intensity Respond to Macroeconomic Variables? Evidence from British Manufacturing, 1970–1986." Manuscript, Harvard University.

Shapiro, Matthew D. 1996. "Macroeconomic Implications of Variation in the Workweek of Capital." *Brookings Papers on Economic Activity*, (2), pp. 79–119.

Solow, Robert M. 1957. "Technological Change and the Aggregate Production Function." *Review of Economics and Statistics*, *39*, pp. 312–20.

Taylor, John B. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy*, *88* (February), pp. 1–23.

______. 1993. "Discretion versus Policy Rules in Practice." *Carnegie Rochester Conference Series on Public Policy*, *39* (December), pp. 195–214.

Varian, Hal. 1984. *Microeconomic Analysis*. New York: W. W. Norton.

Wen, Yi. 1997. "Capacity Utilization under Increasing Returns to Scale." Manuscript.