H I E R

Harvard Institute of Economic Research

Discussion Paper Number 1986

Are Technology Improvements Contractionary?

by

Susanto Basu
John Fernald
Miles Kimball

December 2002

Harvard University
Cambridge, Massachusetts

Comments Encouraged

ARE TECHNOLOGY IMPROVEMENTS CONTRACTIONARY?

Susanto Basu
University of Michigan and NBER

John Fernald
Federal Reserve Bank of Chicago

Miles Kimball
University of Michigan and NBER

Abstract: Yes. We construct a measure of aggregate technology change, controlling for imperfect competition, varying utilization of capital and labor, and aggregation effects. On impact, when technology improves, input use falls sharply, and output may fall slightly. With a lag of several years, inputs return to normal and output rises strongly. We discuss what models could be consistent with this evidence. For example, standard one-sector real-business-cycle models are not, since they generally predict that technology improvements are expansionary, with inputs and (especially) output rising immediately. However, the evidence is consistent with simple sticky-price models, which predict the results we find: When technology improves, input use generally falls in the short run, and output itself may also fall.

First draft: June 21, 1997
This revision: August 2001

We thank Robert Barsky, Menzie Chinn, Russell Cooper, Christopher Foote, Jordi Galí, Dale Henderson, Michael Kiley, Lutz Kilian, Robert King, Jonathan Parker, Shinichi Sakata, Matthew Shapiro, Jonathan Wright and seminar participants at a number of institutions and conferences. Basu and Kimball gratefully acknowledge support from a National Science Foundation grant to the NBER. Basu also thanks the Alfred P. Sloan Foundation for financial support. This paper represents the views of the authors and should not be interpreted as reflecting the views of the Federal Reserve system or other members of its staff.

When technology improves, does employment of capital and labor rise in the short run? Standard frictionless real-business-cycle models generally predict that it does. By contrast, other macroeconomic models predict the opposite. For example, sticky-price models generally predict that technology improvements cause employment to fall in the short run, when prices are fixed, but rise in the long run, when prices change. Surprisingly, plausible sticky-price models also imply that technology improvements may reduce output as well as inputs in the short run. Hence, correlations among technology shocks, inputs, and output shed light on the empirical merits of different business-cycle models.

Measuring these correlations requires an appropriate measure of aggregate technology. We construct such a series by controlling for non-technological effects in the aggregate Solow residual: increasing returns, imperfect competition, varying utilization of capital and labor, and aggregation effects.[1] Our "corrected" technology residual varies about one-half as much as the Solow residual. In addition, though the Solow residual is strongly procyclical, technology fluctuations tend to be countercyclical—contemporaneously, they have a significantly negative correlation with inputs, and a near-zero correlation with output.

We then explore the dynamic response of the economy to technology shocks. Technology improvements reduce employment within the year, but increase employment with a lag of up to two years. Output falls slightly (though not by a statistically significant amount) in the first year, but increases strongly thereafter. Output ultimately increases about as much as technology improves, roughly as one expects.

Correcting for unobserved input utilization is central for understanding the relationship between the procyclical Solow residual and our countercyclical technology residual. Utilization is a form of primary input. Our estimates imply that when technology improves, unobserved utilization as well as observed inputs fall sharply on impact. Both then recover with a lag. In other words, when technology improves utilization falls—so the Solow residual rises less than technology does.

Of course, if technology shocks were the only impulse—and if, as we estimate, these shocks were negatively correlated with the cycle—then after controlling for utilization, we would still be likely to observe a negative (though weakened) correlation between the observed Solow residual and the business cycle. Demand shocks can explain why, by contrast, the observed Solow residual is procyclical. When demand increases, output and inputs—including unobserved utilization—increase as well. We find that shocks other

---

[1] For some of the many recent references on technology and the Solow residual, see, for example, Basu (1996), Basu and Fernald (1997a), Bils and Cho (1994), Burnside (1996), Burnside et al. (1996), and Shapiro (1996).

than technology are much more important at cyclical frequencies, so changes in utilization make the observed Solow residual procyclical.

We identify technology shocks using the tools of Basu and Fernald (1997a) and Basu and Kimball (1997), who in turn build on Solow (1957) and Hall (1990). Basu and Fernald stress the role of sectoral heterogeneity. They argue that for economically plausible reasons—e.g., differences across industries in the degrees of market power—the marginal product of an input may differ across uses. Growth in the aggregate Solow residual then depends on which sectors change input use the most over the business cycle. Basu and Kimball stress the role of variable capital and labor utilization. Their basic insight is that a cost-minimizing firm operates on all margins simultaneously—whether observed or unobserved—ensuring that the marginal benefit of any input equals its marginal cost. As a result, increases in observed inputs can proxy for unobserved changes in utilization. For example, when labor is particularly valuable, firms will work existing employees both longer (increasing observed hours per worker) and harder (increasing unobserved effort).

Together, these two papers imply that to construct an index of aggregate technology change, one should first "purify" sectoral Solow residuals and then aggregate across sectors. Thus, our fundamental identification comes from estimating sectoral production functions, using instruments that we argue are uncorrelated with true technology change.

Galí (1999) and Kiley (1997) have independently used a quite different method to investigate similar issues. Following Blanchard and Quah (1989) and Shapiro and Watson (1988), they identify technology shocks using long-run restrictions in a structural VAR. In particular, Galí and Kiley make the identifying assumption that only technology shocks can affect labor productivity in the long run. Galí examines aggregate data for a number of countries, while Kiley investigates sectoral data for U.S. manufacturing industries. Like us, they find that technology shocks reduce input use.

Shea (1998) proposes yet another method, and also finds results consistent with ours. He measures technology shocks as innovations to R&D spending and patent activity. With a long lag, he finds that process innovations (as opposed to product innovations) raise measured productivity and reduce inputs. As Galí (1998) notes, innovations generally should have their greatest effect when they increase productivity rather than when the original R&D or patent activity takes place. (Of course, forward-looking variables like consumption or stock prices react when the outcome of the R&D investment is known, which may be before measured productivity changes.) Hence, Shea finds, as we do, that when technology rises, inputs fall.[2]

Our approach has two advantages relative to these papers. First, our results do not depend on long-run identifying assumptions that may not hold. For example, Galí's and Kiley's identifying assumption that only technology shocks change long-run labor productivity is not robust to increasing returns or permanent changes in the composition of output—two non-technology shocks that can change long-run labor productivity.[3] Moreover, even if the long-run restriction holds, it produces well-identified shocks and reliable inferences only under strict conditions (see, for example, Faust and Leeper, 1997). Our production-function approach, by contrast, seeks to identify technology shocks directly. Second, we construct a long time series of technology residuals. Shea's data do not allow him to construct a long time series, nor can he investigate results outside of manufacturing.

Nevertheless, the three approaches are best regarded as complements, with distinct identification schemes and strengths. Despite differing data, countries, and methods, the bottom line is that the three very different approaches give similar results.[4] It thus appears we have uncovered a robust stylized fact: technology improvements are contractionary on impact.

What implications do these results have for modeling business cycles? They are clearly inconsistent with standard parameterizations of frictionless RBC models, including the recent attempt by King and Rebelo (1998) to "resuscitate" these models. However, our findings are consistent with the predictions of dynamic general-equilibrium models with sticky prices. Consider the simple case where 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 technology improves. Firms now need less labor to produce this unchanged output, so they lay off workers and reduce hours.[5] Over time, however, prices adjust, the underlying real-business-cycle dynamics take over, and output rises. Relaxing the quantity-theory assumption allows for richer output dynamics: as we discuss in Section V, output can actually fall after a technology improvement, matching the pattern we observe in U.S. data.

Of course, in a sticky-price model, the technology improvements will be contractionary only if the

---

[2] Zachariadis (1999) finds that R&D and patent activity have a statistically significant relationship with our measure of technology change, with substantial lags. The relationship is somewhat weaker with the standard Solow residual.

[3] Sarte (1997) argues that Galí's results are sensitive to alternative "reasonable" long-run identifying assumptions.

[4] Jordi Galí tells us that for the United States, his measure of technology has a correlation of 0.6 with ours.

[5] Tobin (1955) makes essentially this argument, in a model with an exogenously fixed nominal wage.

monetary authority does not offset their short-run effects through expansionary monetary policy. After all, standard sticky-price models predict that a technology improvement that increases full-employment output creates a short-run deflation, which in turn gives the monetary authority room to lower interest rates. In Section V, we argue that technology improvements are still likely to be contractionary, reflecting the fact that central banks react with a lag as well as the fact that inflation may be sluggish, leading the central bank to keep interest rates close to pre-shock levels for a long time.

Clearly, our results are not a "test" of sticky-price models of business cycles, even though the results are consistent with that interpretation. We favor this interpretation in part because sticky-price models are desirable on other grounds, notably, their ability to generate large monetary non-neutralities. Nevertheless, other explanations are possible, including a flexible-price world with autocorrelated technology shocks; sectoral shifts, if reallocations are larger when mean technology is higher; the need to learn about new technologies, leading to unobserved investments in knowledge; and cleansing effects of recessions, in which recessions lead to reorganization within a firm or the elimination of low-productivity firms within an industry.[6] We discuss these alternative explanations at length in Section V. We conclude that the sticky-price explanation seems the most consistent with the data.

The paper has the following structure. Section I reviews our method for identifying sectoral and aggregate technology change. Section II discusses data and econometric method. Section III presents our main empirical results. Section IV discusses robustness. Section V presents alternative interpretations of our results, including our preferred sticky-price interpretation. Section VI concludes.

I. Estimating Aggregate Technology, Controlling for Utilization

This section describes our basic method of identifying aggregate technology. The basic idea is to estimate Hall-style regression equations at a disaggregated level, with proxies for utilization. We then define aggregate technology change as an appropriately-weighted sum of the resulting residuals. Subsection A discusses the augmented Solow-Hall approach and method of aggregation, while Subsection B discusses the theory underlying our method of controlling for utilization.[7]

---

[6] Yorokuglu (1999) argues that with indivisibilities in consumption, process innovations may reduce input use.
[7] Basu and Fernald (1999) provide detailed derivations and discussion of the equations in this section.

A. Firm and Aggregate Technology

We assume each firm has a production function for gross output:

$$Y_i = F^i\left(A_i K_i, E_i H_i N_i, M_i, Z_i\right). \tag{1.1}$$

The firm produces gross output, Y_i, using the capital stock K_i, employees N_i, and intermediate inputs of energy and materials M_i. We assume that the capital stock and number of employees are quasi-fixed, so their levels cannot be changed costlessly. However, firms may vary the intensity with which they use these quasi-fixed inputs: H_i is hours worked per employee; E_i is the effort of each worker; and A_i is the capital utilization rate (i.e., capital's workweek). Total labor input, L_i, is the product $E_iH_iN_i$. The firm's production function F^i is (locally) homogeneous of arbitrary degree γ_i in total inputs. If γ_i exceeds one, then the firm has increasing returns to scale, reflecting overhead costs, decreasing marginal cost, or both. Z_i indexes technology.

Following Hall (1990), we assume cost minimization and relate output growth to the growth rate of inputs. The standard first-order conditions give us the necessary output elasticities, i.e., the weights on growth of each input. Let dx_i be observed input growth, and du_i be unobserved growth in utilization. (For any variable J, we define dj as its logarithmic growth rate $\ln(J_t / J_{t-1})$.) The firm charges a markup μ_i of price over marginal cost, where the markup equals one if the firm is perfectly competitive. We find:

$$dy_i = \mu_i(dx_i + du_i) + dz_i\,, \tag{1.2}$$

where

$$dx_i = \left[s_{Ki}dk_i + s_{Li}(dn_i + dh_i) + s_{Mi}dm_i\right], \tag{1.3}$$

$$du_i = \left[s_{Ki}da_i + s_{Li}\,de_i\right],$$

and s_{Ji} is the ratio of the cost of input J to total revenue. Section *I.B* explores ways to measure du_i.

How are the firm-level technology shocks dz_i, defined (implicitly) by equation (1.2), 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 (i.e., $C + I + G + X - M$), 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, the correct definition is unambiguous, since the macro perspective yields a unique measure that is equivalent to the natural micro measure (Domar, 1961; Hulten, 1978). Rotemberg and Woodford (1995) show that equivalence also holds with imperfectly competitive product markets, but only under certain restrictive conditions: factor markets must be competitive, and all firms must have identical separable gross-output production functions, charge prices that are the same markup over marginal cost, and always use 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 may lead to different definitions; indeed, neither the macro nor the micro perspectives yield a single unambiguous measure. 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 variations as "technology change," since they can occur with no change in the technology available to any firm.

Now consider the following definition: Technical change is 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. In particular, Basu and Fernald (1997b) show that this measure of technical change equals:

$$dz = \sum_i w_i \frac{dz_i}{1-\mu_i s_{Mi}} , \tag{1.4}$$

where w_i equals $\left(P_i Y_i - P_{Mi} M_i\right) \big/ \sum_i \left(P_i Y_i - P_{Mi} M_i\right) \equiv P_i^V V_i \big/ P^V V$, the firm's share of aggregate nominal value added. Conceptually, this measure first converts the gross-output technology shocks to a value-added basis by dividing through by $1-\mu s_M$. (A value-added basis is desirable because of the national accounts identity, which tells us that aggregate final expenditure equals aggregate value added.) These value-added shocks are then weighted by the firm's share of aggregate value added.

Equation (1.4) defines a "micro" measure of technical change, since it changes only if firm-level technology changes. However, it nests the Rotemberg-Woodford definition as a special case, and thus 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. We thus focus on definition (1.4) in constructing aggregate technology.

However, a disadvantage of the measure in equation (1.4) is that it requires the firm-level markups. Domar (1961) and Hulten (1978) propose a different definition of aggregate technology:

$$dz' = \sum_i w_i \frac{dz_i}{1 - s_{Mi}} \tag{1.5}$$

They show that equation (1.5) satisfies both the micro and macro definitions of technical change when there are constant returns and perfect competition; (1.4) then reduces to (1.5).

With imperfect competition, the Domar-weighted measure shows how much final output changes from changes in firm-level technology, holding fixed both the aggregate quantities and the distributions of primary *and* intermediate inputs. We find this definition of aggregate technical change unappealing, since it corresponds to a thought experiment where firms cannot use more intermediate inputs even when they receive favorable technology shocks. However, since this measure has the advantage of not requiring knowledge of sectoral markups, we use it to check the robustness of our primary measure.

We define changes in aggregate utilization as the contribution to final output of changes in firm-level utilization. This, in turn, is a weighted average of firm-level utilization change du_i, estimated using one of the methods in the next sub-section:

$$du = \sum_i w_i \frac{\mu_i du_i}{1 - \mu_i s_{Mi}} \tag{1.6}$$

Note from equation (1.2) that $\mu_i du_i$ enters in a manner parallel to dz_i and hence (1.6) parallels (1.4).

B. Measuring Firm-Level Capacity Utilization

Utilization growth, du_i, is a weighted average of growth in capital utilization, A_i, and labor effort, E_i. The challenge in estimating firm and aggregate technology using equations (1.2) and (1.4) is to relate du_i to observable variables. We do so following Basu and Kimball (1997), who use the basic insight that a cost-minimizing firm operates on all margins simultaneously, so the firm's first-order conditions imply a relationship between observed and unobserved variables. Thus, increases in observed inputs can, in principle, proxy for unobserved changes in utilization.

The model of this section provides theoretical microfoundations for a simple proxy. In particular, changes in hours worked proxy appropriately for unobserved changes in both effort and capital utilization. We are thus able to control appropriately for variable utilization without (perhaps heroically) assuming that one can observe either the firm's internal shadow prices of capital, labor and output; or the true quantities of capital and labor input at high frequencies. Our results use only the cost-minimization problem and the assumption that firms are price-takers in factor markets; we do not require any assumptions about the firm's pricing and output behavior in the goods market.

Following Basu and Kimball (1997), we model the firm 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. We believe that 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 there were no cost to increasing the rate of investment or hiring, 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.[8]

We 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. We assume the major cost of increasing capital utilization, A, is that firms may have to pay a shift premium (a higher base wage) to compensate employees for working at night, or at other undesirable times.[9] We take A to be a continuous variable for simplicity, although variations in the workday of capital (i.e., 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.[10] The cost of higher labor utilization, E, is a higher disutility on the part of workers that must be compensated with a higher wage. We allow for the possibility that high-frequency fluctuations in this wage might be unobserved, as could be the case if wage payments are governed by an implicit contract in a long-term relationship.

We consider the following dynamic problem, in which an industry's representative firm minimizes the

---

[8] One does not require *internal* adjustment costs to model variable factor utilization in an aggregative model (see, e.g., Burnside and Eichenbaum, 1996), since changes in the representative firm's input demand affects the aggregate real wage and interest rate. However, since we want to model the behavior of industries that vary utilization in response to idiosyncratic changes in technology or demand, we require internal adjustment costs in order to have a coherent model of variable factor utilization. (Haavelmo's (1960) treatment of investment makes both of these observations.)

[9] Basu and Kimball (1997) extend this model to allow utilization to affect the rate at which capital depreciates.

[10] See, e.g., Beaulieu and Mattey (1998) and Shapiro (1996).

present value of expected costs:

$$\underset{A,E,H,M,I,R}{\text{Min}} \quad E_t \sum_{s=t}^{\infty} \left[\prod_{j=t}^{s} \left(1+r_j\right)^{-1} \right] \left[WNG\left(H,E\right)V\left(A\right) + P_M M + WN\Psi\left(R/N\right) + P_I K\, J\left(I/K\right) \right] \quad (1.7)$$

subject to

$$\overline{Y} = F(AK, EHN, M, Z) \quad (1.8)$$

$$K_{t+1} = I_t + (1-\delta)K_t \quad (1.9)$$

$$N_{t+1} = N_t + R_t \quad (1.10)$$

In each period, the firm's costs in (1.7) are total payments for labor and materials, and the costs associated with undertaking gross investment *I* and hiring (net of separations) *R*.. $WG(H,E)V(A)$ is total compensation per worker (which may take the form of an implicit contract, and hence not be observed period-by-period). *W* is the base wage; the function *G* specifies how the hourly wage depends on effort, *E*, and the length of the workday, *H*; and *V*(*A*) is the shift premium. P_M is the price of materials. $WN\Psi(R/N)$ is the total cost of changing the number of employees; $P_I K\, J(I/K)$ is the total cost of investment; δ is the rate of depreciation. We omit time subscripts where this practice does not cause confusion.

We assume that Ψ and *J* are convex, and make the appropriate technical assumptions on *G* in the spirit of convexity and normality.[11] We make some normalizations relative to normal or "steady-state" levels of the variables.[12] Let $J(\delta) = \delta,\ J'(\delta) = 1,\ \Psi(0) = 0$. We also assume that the marginal employment adjustment cost is zero at a constant level of employment: $\Psi'(0) = 0.$

There are six intra-temporal first-order conditions and two Euler equations, for the state variables *K* and *N*. To conserve space, we analyze only the optimization conditions that affect our derivation; Basu and Kimball (1997) discuss the full problem in detail. Let λ be the multiplier on constraint (1.8); λ has the interpretation of marginal cost. Numerical subscripts denote derivatives of the production function *F* with respect to its first, second and third arguments, and literal subscripts denote derivatives of the labor cost function *G*. The

---

[11] 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.

[12] If there is a trend, the model can be expressed in terms of detrended quantities from the beginning.

conditions that we require are those for optimization with respect to choices of *A*, *H*, and *E*. These are:

$$A: \qquad \lambda F_1 K = wNG(H,E)V'(A) \tag{1.11}$$

$$H: \qquad \lambda F_2 EN = wNG_H(H,E)V(A) \tag{1.12}$$

$$E: \qquad \lambda F_2 HN = wNG_E(H,E)V(A) \tag{1.13}$$

Note that the firm's uncertainty about future variables does not affect our derivations, which rely only on (some of) the *intra*-temporal equations for optimization. Uncertainty affects the evolution of the state variables (as the Euler equations would show) but not the minimization of variable cost at a point in time, *conditional* on the levels of the state variables. Our utilization proxy depends only on this static variable-cost minimization.

Equations (1.12) and (1.13) can be combined into an equation implicitly relating *E* and *H*:

$$\frac{HG_H(H,E)}{G(H,E)} = \frac{EG_E(H,E)}{G(H,E)}. \tag{1.14}$$

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*, (1.14) implies a unique, upward-sloping *E*-*H* expansion path, so that we can write

$$E = E(H), \qquad E'(H) > 0. \tag{1.15}$$

Equation (1.15) expresses unobserved intensity of labor utilization *E* as a function of the observed number of hours per worker *H*. We define $\zeta \equiv H^* E'(H^*)/E(H^*)$ as the elasticity of effort with respect to hours, evaluated at the steady state. Log-linearizing, we can write the growth rate of effective labor input as:

$$d\ln(EHN) = dn + dh + de = dn + (1+\zeta)dh. \tag{1.16}$$

To find a proxy for capital utilization, we combine (1.11) and (1.12). Rearranging, we find:

$$\frac{F_1 AK/F}{F_2 EHN/F} = \left[\frac{G(H,E)}{HG_H(H,E)}\right]\left[\frac{AV'(A)}{V(A)}\right] \tag{1.17}$$

The left-hand side is a ratio of output elasticities, which (as in Hall 1990) one can show are proportional to factor cost shares when cost is minimized. We denote these cost shares by α_K and α_L. Define *g(H)* as the elasticity of cost with respect to hours, and $v(A)$ as the ratio of the marginal to the average shift premium:

$$g(H) = \frac{HG_H\big(H, E(H)\big)}{G\big(H, E(H)\big)} \tag{1.18}$$

$$v(A) = \frac{AV'(A)}{V(A)}. \tag{1.19}$$

With these definitions, we can write equation (1.17) as:

$$v(A) = \frac{\alpha_k}{\alpha_L} g(H). \tag{1.20}$$

The labor cost elasticity with respect to hours given by the function $g(H)$ is positive and increasing by the assumptions we have made on $G(H,E)$. The labor cost elasticity with respect to capital utilization, given by the function $v(A)$, is positive as long as there is a positive shift premium. We assume that the shift premium increases rapidly enough with *A* to make the elasticity increasing in *A*. We also assume that α_K/α_L is constant, which requires that *F* be a generalized Cobb-Douglas in *K* and *L*.[13]

Under this assumption, the log-linearization of (1.19) is simply

$$da = \frac{\eta}{\omega} dh. \tag{1.21}$$

where η is the elasticity of *g* with respect to *H* and ω is the elasticity of *v* with respect to *A*. η indicates the rate at which the elasticity of labor costs with respect to hours increases. ω indicates the rate at which the elasticity of labor costs with respect to capital utilization increases.

Thus, equations (1.21) and (1.15) say that the change in hours per worker should be a proxy for changes in *both* unobservable labor effort and the unmeasured workweek of capital. The reason that hours per worker proxies for capital utilization as well as labor effort is that shift premia create a link between capital hours and labor compensation. The shift premium is most worth paying when the marginal hourly cost of labor is high relative to its average cost, which is the time when hours per worker are also high.

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

---

[13] Thus, we assume $Y = Z\Gamma\big((AK)^{\alpha_K}(EHN)^{\alpha_L}, M\big)$, where Γ is a monotonically increasing function. As Basu and Kimball (1997) note, estimating the general case where the ratio of the elasticities is a function of all four input quantities would demand too much of the data and the instruments.

$$dy = \mu\, dx + \mu\left(\zeta s_L + \frac{\eta}{\omega} s_K\right) dh + dz$$
$$= \mu\, dx + \beta dh + dz. \tag{1.22}$$

We will not need to identify all of the parameters in the coefficient multiplying dh, so we denote that composite coefficient by β. This specification controls for *both* labor and capital utilization.[14]

So far we have discussed only model-based proxies, using cost-minimizing conditions derived under fairly general assumptions. One strand of literature controls for variable utilization under more restrictive assumptions of fixed proportions between an observed and unobserved input. For example, Burnside et al. (1995, 1996) revive the suggestion of Jorgenson and Griliches (1967) and Flux (1913) that electricity use is a natural proxy for total capital services. As a check on our model-based proxies, we assume:

$$da + dk = d(electricity). \tag{1.23}$$

This procedure ignores variations in labor utilization. It is also more appropriate for heavy equipment than structures, and hence may be a good proxy for capital input only in manufacturing industries.

II. Data and Method

A. Data

We now construct a measure of "true" aggregate technology change, *dz*, and explore its properties. As discussed in the previous section, we estimate technology change at a disaggregated level, and then aggregate. Our aggregate is the private non-farm, non-mining U.S. economy.

Since the theory applies to firms, it would be preferable to use firm-level data. Unfortunately, no firm-level data sets span the economy. In principle, we 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, our paper complements existing work that uses small subsets of the economy.

We use data compiled by Dale Jorgenson and Barbara Fraumeni on industry-level inputs and outputs.

---

[14] Basu and Kimball (1997) generalize this model to allow depreciation to vary depending on capital utilization, as in a variety of papers. This modification introduces two new terms into the estimating equation, but Basu and Kimball cannot reject the hypothesis that these terms are insignificant; in any case, including them leads to results that are virtually identical to those reported below.

The data comprise 29 industries (including 21 manufacturing industries at roughly the two-digit level) that cover the entire non-farm, non-mining 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. (For a complete description of the dataset, see Jorgenson et al., 1987.) Our data run from 1947 to 1989; in our empirical work, however, we restrict our sample to 1950 to 1989, since our money shock instrument is not available for previous years.

We compute capital's share s_K for each industry by constructing a series for required payments to capital. We 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, we use data on the current value of the 51 types of capital, plus land and inventories. For each of these 53 assets, indexed by s, the user cost of capital R_s is $(r+\delta_s)(1-ITC_s-\tau d_s)/(1-\tau)$. *r* is the required rate of return on capital, 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. We 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.

Our empirical work requires instruments uncorrelated with technology change. We 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. (We use the contemporaneous value and one lag of each instrument.) We also use a version of the instrument used by Burnside (1996): quarterly Federal Reserve "monetary shocks" from an identified VAR. We sum the four quarterly policy shocks in year *t-1* as instruments for year *t*.[15]

B. Estimating Technology Change

To estimate "firm-level" technology change, we take the residuals from industry regressions of (1.22). We expect that residuals may be correlated across industries, so there are efficiency gains from estimating

---

[15] We drop the third Hall instrument, the political party of the President, because it has little relevance in any industry. Burnside (1996) argues that the oil price instrument is generally quite relevant, and defense spending explains a sizeable fraction of input changes in durable-goods. The qualitative features of the results in Section III are robust to different combinations and lags of the instruments. Section IV considers the small sample properties of instrumental variables.

these regressions as a system of equations. One concern is that the utilization coefficients are often estimated rather imprecisely. To mitigate this problem, we combine industries into three groups on *a priori* grounds, and restrict the utilization parameters to be constant within these groups. Thus, for each group we have

$$dy_i = c_i + \mu_i dx_i + \beta dh_i + dz_i \,. \qquad (2.1)$$

This parsimonious equation allows us to control for both capital and labor utilization if the cost of higher capital utilization is a shift premium. The markup μ_i differs by industries within a group (Burnside (1996) emphasizes the importance of allowing this variation). The groups are durables manufacturing (11 industries); non-durables manufacturing (10); and all others, such as construction, services, and utilities (8). We also estimate this regression omitting the hours-per-worker term to obtain residuals that do not correct for utilization. Industry technology change is then the sum of the industry-specific constant c_i and residual dz_i. To avoid the "transmission problem" of correlation between technology shocks and input use, we estimate each system using three-stage least squares, using the instruments noted in Section *II.A.*

We confirmed that results are robust to using industry-by-industry rather than group estimation. We estimated the individual equations with both two-stage-least-squares and LIML (which Staiger and Stock (1997) argue is less subject to small-sample biases). Parameters are more variable with individual than group estimation, particularly the LIML estimates, but the median estimates are in both cases similar to the median 3SLS estimates. Estimating individual equations substantially increases the variance of the estimated aggregate technology residuals, but the main correlation results below are not qualitatively affected.[16]

III. Results

A. Basic Correlations

We quickly summarize the parameter estimates from our Hall-style industry regressions. The main focus of this and the sections which follow are the aggregate effects of technology shocks, where aggregate technology is estimated as an appropriately weighted average of the industry regression residuals.

---

[16] The higher variance reflects the convexity of (1.4) with respect to the markup μ. Suppose markup estimates are unbiased, but we increase the variance of the estimate around the true value. The convexity of (1.4) then makes dz more sensitive to fluctuations in dz_i (The most extreme case is where the estimate of μs_M is close to one, so that $1/(1-\mu s_M)$ approaches infinity.) This potential sensitivity to estimates of the markup is one reason we look at the Domar-weighted aggregate from equation (1.5); although it has less theoretical basis than (1.4), markup estimates do not affect it. The Domar-weighted residuals turn out to have a correlation of 0.94 with the markup-corrected residuals.

Table 1 shows the parameter estimates from equation (2.1). For the 29 industries, the value-added-weighted average markup estimate is 0.92. For durables, the average is 1.05; for non-durables, 0.87; for non-manufacturing, 0.89. (Omitting the hours-per-worker term, the average markup in durables rises to 1.10, but the overall average rises only to 0.93.) The estimates for non-manufacturing are the least precisely estimated, and the most variable. Fortunately, the contractionary effects of technology improvements presented below are driven primarily by manufacturing, so the variability of non-manufacturing does not seem to be a major problem. Our results are also not driven by any single industry—omitting industries that look like "outliers," for example, has relatively little effect on results.

The coefficient on hours-per-worker, in the bottom panel, is strongly statistically significant in durables manufacturing. In non-durables, it has a t-statistic of 1.6, with a p-value of 0.11. In non-manufacturing, it is completely insignificant. The results below are virtually unaffected if we use the residuals from the regression omitting the hours-per-worker term outside of durables.

Table 2 reports summary statistics for four series. The first is the standard Solow residual, calculated using aggregate data alone, with no adjustments for utilization or markups. The other three measures are derived from sectoral regression residuals. The second makes no corrections for utilization, and is aggregated using our theoretically preferred "markup-weighting" from equation (1.4). The third, our "preferred" series, uses hours-per-worker to control for utilization, and again is aggregated using markup-weighting (1.4). The final series again corrects for utilization, but then uses Domar aggregation from equation (1.5).

Panel A shows results for the entire private non-mining economy. The variance of our corrected series are substantially smaller than for the Solow residual: The variance of the hours-corrected markup-weighted residual is less than half that of the Solow residual, and the standard deviation (shown in the second column) is only about two-thirds as large. We do estimate negative technical change in some periods—see the third column—but the lower variance of the technology series implies that the probability of negative residuals is much lower. The Solow residual is negative in 9 years out of 40; by contrast, the markup-weighted hours-corrected measure (3) is negative in 5 years and the Domar-weighted measure (4) is negative in 3 years.

Panel B gives results within manufacturing alone. Data within manufacturing (especially for output) are often more reliable than data outside manufacturing. In addition, some other papers (such as Burnside et al., 1996) focus only on manufacturing, so these results provide a basis for comparison. The results are qualitatively similar to those for the aggregate economy.

Some simple plots summarize the comovement in our data. Figure 1 plots business-cycle data for the private economy: output (value-added) growth *dv*, primary input growth, dx^V, and the Solow residual *dp* (all series are demeaned). These series comove positively, quite strongly so in the case of *dp* and *dv*.

Figure 2 plots our preferred (markup-weighted hours-corrected) technology series against these three variables. The top panel shows that technology fluctuates much less than the Solow residual, consistent with intuition that non-technological factors, such as variable input utilization, increase the volatility of the Solow residual. In addition, some periods show a phase shift: the Solow residual lags technology change by one to two years. This phase shift reflects the utilization correction. In our estimates, technology improvements are associated with low levels of utilization—reflected in low hours-per-worker, which in our model implies low unobserved effort—thereby reducing the Solow residual relative to the technology series. Hours per worker generally increase strongly a year after a technology improvement, raising the Solow residual.

The middle panel plots aggregate value-added output growth (*dv*) against technology. There is no clear contemporaneous comovement between the two series—although, again, the series appear to have a phase shift: output comoves with technology, lagged one to two years.

Finally, the bottom panel plots the growth rate of primary inputs of capital and labor (dx^V) and the same technology series. These two series clearly comove negatively over the entire sample period.

The comovements between technology and input and output are clearly inconsistent with those found in the usual RBC literature. By contrast, in Section V, we suggest a sticky-price model that is consistent with Figure 2. In that model, the contemporaneous correlation between technology shocks and inputs is negative; the contemporaneous correlation of output growth and technology shocks is ambiguous. Correlations turn positive with a lag, thus explaining the apparent phase shift in the figures.

We now examine the relationship between technology and other aggregate variables more formally. Table 3 shows simple correlations. We include growth of total hours worked ($dh + dn$). Panel A shows results for the aggregate private economy, and panel B shows results for manufacturing alone.

The top row of either panel shows the usual business-cycle facts: The Solow residual is significantly correlated with output, total inputs, and hours. Hours correlate more strongly with productivity than do total inputs, reflecting the low correlation of changes in the capital stock with the business cycle.

The third and fourth rows of both panels contain the key results of the paper: the correlations between technology and business-cycle variables. These correlations differ sharply from those predicted by the usual

RBC model (e.g., Cooley and Prescott, 1995). For our preferred measure in row 3, the correlation of technology with output is about zero, and the correlations with inputs are strongly negative: -0.37 for total primary inputs, and -0.43 for hours alone. Both correlations are statistically significantly negative at the 95 percent level. The correlation of the corrected series with the Solow residual is positive, at 0.42.

The non-utilization-corrected technology series (row 2) shows the same general tendencies, but to a lesser extent. For example, though the correlation of this series with output is strongly positive at 0.46, it is statistically smaller (at the 90 percent level) than the correlation between the standard Solow residual and output. Non-utilization-corrected technology is not significantly correlated with inputs.

Note that the correlations between all three technology measures and output are statistically smaller than the correlation between the Solow residual and output, at the 90 percent level or better. The correlations between the two utilization-corrected technology measures and inputs are statistically smaller than the correlation between the Solow residual and inputs, at the 95 percent level.[17]

As the top and bottom panels of Table 3 suggest, the contractionary effect of technology on inputs is particularly strong in manufacturing. Indeed, durables manufacturing to a large extent drives our results. For example, using the utilization-corrected series within durables and the non-utilization-corrected series elsewhere leaves the correlation with inputs almost unchanged, -0.48 compared with –0.49. Within manufacturing, the strong negative correlation reflects the adjustments for both utilization and markups.

B. Dynamic Responses to Technology Improvement

Impulse responses to innovations in our technology series provide a simple and convenient way to show dynamic correlations between technology innovations and our basic variables. The aggregate variables we examine are output growth (dv), input growth (dx^V), total hours worked ($dh + dn$), and our constructed series for utilization change, $d\hat{u}$, defined by equation (1.6). We first estimate an AR(2) process for our estimated dz series in order to derive the innovations, ε:

$$d\hat{z}_t = \alpha_0 + \alpha_1 d\hat{z}_{t-1} + \alpha_2 d\hat{z}_{t-2} + \varepsilon_t \,. \qquad (3.1)$$

For dz we use our utilization-corrected markup-weighted measure of technology change; results are virtually identical using the Domar-weighted series. To derive the impulse response of any variable J to a technology

[17] To calculate the t-statistic for the difference in correlations, we assume the two correlations are independent. This is obviously not the case, since technology affects productivity. Taking account of this positive covariance would

innovation, we compute $dj = \sum_{i \geq 0} \phi_i \hat{\varepsilon}_{t-i}$. In practice, to estimate $\hat{\varepsilon}_t$ and the moving-average terms ϕ_i, we estimate equation (3.1), along with a second equation in which we regress *dj* on its own lags and current and lagged values of *dz*. The impulses responses are derived using the estimated system, which we estimate by SUR. In all cases, we use a lag length of two periods (in our case, years).[18]

Note that our procedure amounts to assuming that the technology series is completely exogenous, which is stronger than the standard ordering assumption in a VAR. Using that ordering assumption would amount to including lagged values of *dj* in equation (3.1). Doing so affects our results only slightly. A deeper question is whether the exogeneity assumption is warranted. As a check, we perform Granger causality tests, using a number of plausible variables (e.g., *dv*, dx^V, *dh*, etc.) In all cases, we cannot reject the hypothesis that the technology series is exogenous.

Figure 3 shows the impulse responses to a technology improvement: the effects of a 1 percent (that is, 0.01) technology improvement on the (log) levels of technology, output, inputs, manhours, and utilization. We also present 95 percent confidence intervals, using the RATS Monte Carlo procedure.[19]

The technology series is approximately an AR(1) in first differences. After a one percent innovation, technology increases about another 0.4 percent the following year, then levels off.

Both output and inputs fall on impact: the fall in inputs is strongly significant, regardless of the type of input considered (capital and labor inputs dx^V, manhours, or utilization). The output decline is not significant. Output grows strongly after the shock: two years out, the impulse response differs significantly from zero, with output rising about 1.8 percent. The point estimate shows inputs growing more slowly. dx^V falls 0.8 percent on impact, and then recovers to its pre-shock level (normalized to zero) in three years. (The 95 percent confidence interval is fairly wide three years out, running from 1 percent to -1 percent.)

The results for utilization explain the phase-shift from Figure 2. On impact, technology improvements reduce utilization. The Solow residual depends (in part) on technology plus the change in utilization (see equation (1.2)); the technology improvement raises the Solow residual, but the fall in utilization reduces it.

---

strengthen our argument, since it means that we *overstate* the variance, and hence understate the t-statistics.

18 We do not use cointegration techniques, because levels of output and inputs need not be cointegrated with technology. For example, changes in demographic structure (e.g., the Baby Boom) or in immigration policy can cause permanent changes in the size of the labor force that are not related to technology.

19 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*. We discuss the generated regressor issue in Section V.

Hence, on impact the Solow residual rises less than the full increase in technology. With a lag, utilization increases, which in turn raises the Solow residual relative to technology.

IV. Robustness checks

We now address robustness. We study the properties of technology shocks at the sectoral level; present an alternate method of controlling for utilization (electricity use); consider classical measurement error; and consider the small sample properties of instrumental variables. Our finding of a negative contemporaneous correlation between technology and inputs survives these considerations.

A. Within-Sector Results

We now examine results at a one- and two-digit sectoral level. The sectoral results make it clear that our results are not simply a consequence of our aggregation method. Table 6 present results for 9 (approximately one-digit) industries, as well as average correlations for the 29 industries in our sample. We concentrate on gross-output results, since gross output gives a clearer picture of the pattern of production at the industry level. (Value-added results are generally quite similar.)

Overall, the results are qualitatively similar to the aggregate results in Table 3. The average industry correlation of inputs with the Solow residual (*dp*) is 0.17; the correlation with our fully-adjusted technology residual (*dt)* falls to about -0.10. Our corrections also reduce correlations between output and technology by more than a factor of two: the average correlation falls from 0.57 to 0.25. In regressions not shown, we regressed industry input growth on industry technology residuals; we used a system of seemingly unrelated regressions, with coefficients on current and lagged sectoral technology innovations constrained to be the same across industries. Again, we find that sectoral shocks reduce inputs sharply on impact. For manufacturing industries, the contemporary coefficient is -0.44, with a t-statistic of 20; for non-manufacturing industries, the contemporary coefficient is -0.33, with a t-statistic of 11.

Quantitatively, the sectoral results are less dramatic than the aggregate results, but that is not surprising. After all, we expect average industry correlations to be smaller than the aggregate correlation, for the simple reason that idiosyncratic shocks increase sectoral standard deviations in the denominator. The aggregate and sectoral results may differ for two other reasons as well. First, the economic effects on a sector from a common, widespread technology improvement may differ substantially from the effects of sector-specific shock. After all, the general equilibrium consequences of a common shock are much larger. For example, economy-wide technology improvements appear deflationary, which in turn tends to be contractionary. At the same time, aggregate shocks have greater wealth effects.

Second, from a mechanical perspective, sectors are not weighted equally in the aggregate results: sectors with large cyclical fluctuations have disproportionate weight. In particular, cyclical fluctuations are very large in construction and durable-goods manufacturing, industries where our corrections are extremely important. For example, our corrections reduce the output and input correlations for manufacturing durables from 0.76 and 0.66 (both significantly positive) to -0.56 and -0.61 (both significantly negative). This pattern makes sense. As Stigler (1939) suggests, industries where demand shocks are frequent and large may build flexibility—including scope for varying factor utilization—into the production process. Thus, our corrections matter disproportionately for the very industries that account for a disproportionate share of fluctuations in output and input use, and thus are more important in producing the aggregate results.

Kiley (1998) correlates our industry residuals within manufacturing with those he derives using Galí's (1999) identification scheme.[20] All but three of the 20 correlations are positive; seven of the positive correlations are significant at the five percent level.

B. Electricity Use as an Utilization Proxy

Electricity use is sometimes used as a proxy for capital utilization. As noted in equation (1.12), we set $du + dk = d(electricity)$ and assume that unobserved effort is constant, so $de = 0$. (Jorgenson and Griliches (1967) and Burnside et al. (1995) agree that electricity proxies only for capital utilization.) Table 5 shows results at the aggregate and manufacturing levels. The top panel uses regression residuals, as described in Section I; the bottom panel uses "corrected" sectoral Solow residuals, as in Burnside et al. (1996), which amounts to setting all sectoral markups to one.[21] The electricity proxy reduces the input and output correlations less than our other corrections. For example, technology's correlation with output is generally around 0.4; in Table 3, it is close to zero or negative. Only in manufacturing, with regression residuals, can we not reject the hypothesis of a zero correlation. Nevertheless, the input correlations—the main focus of our paper—are always close to zero or negative, consistent with the results using our model-based proxies.

Although qualitatively similar, our results differ quantitatively from those of Burnside et al. (1996),

---

[20] Kiley's work is reported in his (1997) paper.

[21] Following Burnside et al., output is value added and inputs are a weighted average of capital and manhours.

who find that the electricity correction reduces technology's correlation with output to 0.1 in manufacturing. Different output data appear to explain the difference in results. Burnside et. al (1996) use monthly industrial production as their basic output measure, averaged to either quarterly or annual frequency. These data have two shortcomings. First, in about one-third of manufacturing industries, IP is calculated from electricity usage. Thus, subtracting electricity use from this measure of "output" biases results towards finding no correlation between technology and output. Second, since Burnside et. al do not have data on intermediate inputs, gross output data (such as IP) is only appropriate under strong separability assumptions, as they note.

On *a priori* grounds, electricity use probably cannot, in general, proxy for utilization. At best, it proxies only for capital utilization, and even there it is most appropriate only for heavy equipment. Variations in utilization of heavy equipment probably occurs mostly in manufacturing, not in, say, Finance or Services—two industries together producing twice the value added of manufacturing. However, for the manufacturing sector alone, results based on electricity-corrected residuals agree fairly closely with ours.

C. Classical Measurement Error

In principle, input measurement error could impart a negative bias to the correlation between technology and inputs. We explore this issue with a simple model, which suggests that measurement error cannot explain our results. First, for plausible parameterizations of the importance of measurement error, the "true" correlation remains negative. Second, the observed covariance between measured output and technology, which is zero or negative, bounds the covariance between true technology and true inputs, again suggesting a negative "true" correlation.

In our empirical work, we take the entire regression residual as "technology," which implicitly assumes that our utilization proxies control fully for all variations in utilization. If they do not, but merely provide unbiased estimates of utilization, then the residual includes non-technological "noise" that is completely analogous to classical measurement error. Our model here abstracts from variations in utilization and does not explicitly consider aggregation across industries; neither changes the basic message .

Suppose the true economic model is given by

$$dy^* = \mu dx^* + dz^*, \tag{4.1}$$

where the starred variables are unobserved, true values. Both output and inputs are measured with error:

$$dy = dy^* + \eta \tag{4.2}$$

$$dx = dx^* + \varepsilon \,, \tag{4.3}$$

where η and ε are iid, mean-zero variables with variances σ_η^2 and σ_ε^2, respectively. Note that the estimated variances of *dy* and *dx* always exceed their true values: $\sigma_{dx}^2 = \sigma_{dx^*}^2 + \sigma_\varepsilon^2$ and $\sigma_{dy}^2 = \sigma_{dy^*}^2 + \sigma_\eta^2$.

Now suppose we estimate (4.1) by instrumental variables. If the instruments are uncorrelated with the measurement error, then the estimate of μ is consistent. Hence, in the limit, the only source of error in our estimate of technology change is the measurement error in *dy* and *dx*:

$$dz = dz^* + \eta - \mu\varepsilon \,. \tag{4.4}$$

Abstracting from estimation error in μ, equation (4.4) implies that $\sigma_{dz}^2 = \sigma_{dz*}^2 + \sigma_\eta^2 + \mu^2\sigma_\varepsilon^2$. Using equation (4.4), the covariances of estimated technology change with output and input growth are:

$$\operatorname{cov}(dz, dy) = \operatorname{cov}\left(dz^*, dy^*\right) + \sigma_\eta^2 \tag{4.5}$$

$$\operatorname{cov}(dz, dx) = \operatorname{cov}\left(dz^*, dx^*\right) - \mu\sigma_\varepsilon^2 \,. \tag{4.6}$$

In terms of our estimated correlations and regression results, measurement error hence biases up both the estimated covariance between output and technology, and the estimated standard deviation of technology. If the true correlation between output growth and technology change is positive, then the estimated correlation may be biased either towards or away from zero, but cannot turn negative. However, suppose the true correlation between output growth and technology change is negative. Then the estimated correlation is unambiguously biased up (towards zero). Thus, our point estimates in Table 3 of a negative correlation between output growth and technology change (statistically significant for manufacturing) cannot be attributed to measurement error.

On the other hand, if the true correlation between input use and technology change is positive, then the estimated input correlation is biased down. If the true input correlation is negative, the estimated correlation may be biased up or down. To get a sense of the potential magnitude of the bias from input measurement error, we write (4.6) in terms of correlations rather than covariances. Some algebra yields:

$$Corr(dz^*, dx^*) = \left[Corr(dz, dx) + \mu \frac{\sigma_\varepsilon^2}{\sigma_{dz}\sigma_{dx}} \right] \left[\frac{\sigma_{dz}\sigma_{dx}}{\sigma_{dz^*}\sigma_{dx^*}} \right]$$

By specifying a markup and a variance for ε,we can calibrate this equation to the observed correlations and variances. A markup of 1, for example, and a standard deviation of measurement error σ_ε of ½ percent per year (compared with observed σ_{dx} of 1.7 percent per year) would imply a "true" correlation of –0.43, compared with the observed correlation of –0.49. This parameterization is close to the maximum permissible standard deviation for σ_{ε}, or else the data would violate a boundedness condition that we now discuss. That boundedness condition gives another indication that measurement error does not drive our results.

We are mostly interested in the signs of the correlations rather than their sizes. We can use the upward-biased output covariance to bound the input-covariance from above. Equation (4.1) implies that

$$\operatorname{cov}\left(dz^{*}, dy^{*}\right) \geq \operatorname{cov}\left(dz^{*}, dx^{*}\right), \tag{4.7}$$

(since the variance of dz^{*} is positive and $\mu \geq 1$). But we see from equation (4.5) that

$$\operatorname{cov}(dz, dy) \geq \operatorname{cov}\left(dz^{*}, dy^{*}\right).$$

Since the estimated covariance of output and technology is either zero or negative, we conclude that the true covariance of technology and inputs must also be zero or smaller. Thus, our surprising results about the effects of technology improvements survive considerations of measurement error.

Since we cannot observe measurement error directly, we cannot say how much it affects our results. However, since the bias works against our finding that technology improvements reduce output, it seems likely that technology improvements are in fact contractionary. Furthermore, unlike the simple model used here, our technology change series takes a weighted average of technology shocks across sectors. If measurement error is relatively independent across industries, averaging should attenuate any biases.

D. Small Sample Properties of Instrumental Variables

We now present several informal tests that indicate that the small sample properties of instrumental variables are probably not having a major effect on our results. For the results presented in Section III, the first-stage regression of industry inputs *dx* on the instruments typically has an F statistic of around 4—statistically significant, but lower than one would like. Staiger and Stock (1997), for example, find that in some cases, instrumental variables estimators have poor small sample properties when the first-stage F statistic is less than about 10.[22]

---

[22] On a priori grounds, all of the variables have strong grounds for being included, and have some relevance for at least some industries. The instruments we use have close to the best overall fit (measured by mean and median F statistic) of the a priori plausible combinations we considered (measures with better "overall" fit generally perform very poorly in a couple of industries). Of course, Hall, Rudebusch and Wilcox (1996) argue that with weak instruments, one does not

First, Staiger and Stock note that LIML has better small sample properties than TSLS. As we noted in Section II, LIML gives results that are qualitatively similar, though more variable, than our preferred results.

Second, we pooled industries within groups in order to raise the significance level of the first stage regression. To implement this, we stacked industries and then estimated equation (2.2) as a stacked regression. This approach raises various econometric issues, including the following. (i) To deal with differences in means (fixed effects), we demeaned all variables before stacking them. (ii) To address concerns about possible heteroskedasticity, we estimated equation (2.2) by OLS for each industry, and then divided all variables by the regression standard error. (iii) Ideally, the first-stage coefficients would be the same across industries within a group—otherwise, the first-stage fit need not improve. Indeed, in non-manufacturing this appeared to be a major problem, since the *signs* of coefficients were not always the same; in some industries, for example, input growth and oil prices are *positively* correlated, while in most industries the correlation is negative.[23] Given that our results appeared to be driven by manufacturing, we resolve this issue by limiting our groups to durables and non-durables. For both durables and non-durables, the first-stage regressions of input growth and hours-growth on the instruments have large F statistics of between 10 and 14—corresponding to p-values of less than 1/10 of 1 percent.

After estimating the pooled regressions, we "unmake" the residuals into industry residuals, and aggregate as before. The resulting technology series has a correlation of 0.8 with our "preferred" manufacturing series (row 3) of Tables 2 and 3. The correlation between technology and output is about -0.27, considerably less negative than the results in Table 3, but nevertheless statistically significant at the 10 percent level, and with relatively similar dynamics. (It is not surprising that the correlation is less negative, given that we lose some of the "reallocation" effects that come from allowing for differences in markups.)

As a final specification test, we simulated 100 draws of random, irrelevant instruments, in the spirit of Bound, Jaeger, and Baker (1995), to get a sense of the actual small sample distribution of IV. From the literature, one expects a bias towards the OLS estimates, which yield a small positive correlation between the "technology residual" and inputs. Indeed, for five simulated instruments, the median simulated correlation for the markup-weighted estimate is –0.02; for the Domar-weighted series, the median is 0.15. Only 3 out of

---

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.

[23] Pooling in non-manufacturing gave a very low first-stage fit for the instruments, and unreasonable parameter values. If we "peek" at the first-stage industry regressions—a dubious practice—and pool into two groups for non-manufacturing

100 simulated correlations for the markup-weighted measure were as negative as our observed correlation;. The simulated correlations for the Domar-weighted measure was rarely negative, and only once was as negative as our observed correlation. It thus appears highly unlikely that the negative correlation reflects small sample biases of weak instrumental variables.

V. Interpretations of the Results

A. The Standard RBC Model

Are the responses of inputs and output to a technology improvement consistent with the standard RBC model (e.g., Cooley and Prescott, 1995)? The long-run impact is certainly consistent. Consider the Cooley-Prescott model with a capital share of 0.35. If Hicks-neutral technology rises 1.4 percent (one standard deviation), then in the long run output rises about 2.2 percent (i.e., 1.4/(1 - 0.35)), inputs (including capital) rise about 0.8 percent, and manhours and utilization do not change. The point estimate for the output response is 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 manhours and utilization.

Our short-run findings, however, are strikingly at odds with the RBC model, at least as calibrated by Cooley and Prescott. In their model, when technology improves, output, investment, consumption, and labor hours all rise on impact. By contrast, the data strongly indicate that technology improvements reduce labor input. Nevertheless, alternative calibrations of the model could deliver this result. Technology improvements raise real wages, which has both income and substitution effects. If the income effect dominates, labor input will fall. Of course, positively autocorrelated technology change would reinforce this fall, since then both income and substitution effects tend to push towards lower current labor supply: workers might take more leisure initially, and work harder in the future, when technology is even better. Indeed, our technology process is positively autocorrelated. But if the leisure story were correct, employment should increase sharply when technology reaches its maximum, but a year after impact inputs are still significantly lower than their pre-shock value.[24]

Nevertheless, output's response most strongly contradicts basic RBC theory. Our point estimates show that output falls, although the response is not significantly different from zero. That technology

---

based on signs of the parameters, we get "better" results, which turn out to look similar to those reported below.

[24] We discuss the issue of autocorrelated technology improvements further in the next subsection.

improvements reduce output as well as inputs on impact contradicts standard flexible-price RBC models driven by permanent technology shocks, since one can show that if consumption and leisure are both normal goods, then an improvement in technology must increase output (see Kimball, 1999).

The data also contradict King and Rebelo's (1998) attempt to "resuscitate" the RBC model. They add variable utilization of capital to the basic RBC model, substantially improving the model's ability to propagate shocks. They use their calibrated model to back out an implied technology series from the Solow residual. By construction, their procyclical technology series, however small, drives business cycles. Our empirical work, by contrast, does not impose such a tightly specified model—and the data reject the King and Rebelo model. Hence, their model does not provide a viable explanation of business cycles any more than the basic RBC model does. Instead, the main lesson we take from their paper is the importance of utilization as a propagation mechanism, a lesson which applies to more realistic models as well.

Thus, the short-run effects of technology improvements contrast sharply with the predictions of standard RBC models. However, are those models right in assuming that technology shocks are the dominant source of short-run volatility of output and inputs? Table 4 reports variance decompositions from the impulse responses in Figure 3. At the business-cycle frequency of three years, technology shocks account for about one-third of the variance of output, but only 10-20 percent of the variance of different input measures. The patterns are intuitively sensible: manhours and utilization respond much more to technology at high frequencies. (Steady-state growth, of course, requires that long-run labor supply be independent of the level of technology.) By contrast, technology accounts for only about 5 percent of the short-run variance of the Solow residual, but almost 60 percent with a lag of three years. Again, this pattern accords with our priors: in the short run, changes in utilization and composition account for much of the volatility of measured productivity. But in the long run, as we expect, the Solow residual reflects primarily changes in technology. Our findings thus lie between the positions of the RBC and New Keynesian schools. Technology shocks are neither the main cause of cyclical fluctuations, nor negligible. Future models should allow for technology shocks, while making sure that the impulse responses of a model match those that we and others find.

B. Price Stickiness

Technology improvements can easily be contractionary in a sticky-price model, in contrast to the one-sector flexible-price RBC model. Suppose the quantity theory governs the demand for money and that the supply of money is fixed. If prices cannot change in the short run, then neither can real balances or output.

Now suppose technology improves. Since the price level is sticky and demand depends on real balances, output does not change in the short run. But firms need fewer workers to produce this unchanged output, so they lay off workers and reduce hours. Over time, however, as prices fall, the underlying RBC dynamics take over. Output rises, and the higher marginal product of capital stimulates capital accumulation. Work hours eventually return to their steady state level.

If money demand depends on interest rates as well as output, then a technology improvement could even cause output to fall in the short run. Output falls if the technology shock causes an excess demand for money at the original level of output. For example, if the money supply is unchanged, a technology improvement reduces expected inflation, since the price level must fall over time. At a given real interest rate, lower inflation raises money demand. Output then must fall to restore money-market equilibrium. In addition, because improved technology allows the same output to be produced with less capital and labor, it will tend to reduce investment demand in the short run, which also tends to cause output to fall.

These effects are present in virtually any dynamic general-equilibrium model with sticky prices, such as Kimball's (1998) Neomonetarist model. Basu (1998) calibrates a DGE model with staggered price setting, and reproduces quite accurately the impact effect of technology improvements that we find in the data.

Of course, the monetary authority is likely to follow a more realistic feedback rule than simply keeping the nominal money stock constant, as our discussion has assumed. Would it accommodate technology improvements by loosening policy, thereby avoiding the initial contraction? Basu (1998) allows the monetary authority to follow a Taylor rule, setting the nominal interest rate in response to lagged inflation and the lagged "output gap"—the deviation between current and full-employment output. He still finds that on impact, output barely changes when technology improves, while inputs fall sharply. Monetary policy is insufficiently loose under a Taylor rule in part because the Federal Reserve reacts only with a lag—that is, after the shock affects inflation or the measured output gap.[25]

We now explore additional evidence to see whether the contractionary effects of technology

---

[25] In Basu's model, the contraction is relatively short-lived, unlike the response we find in the data. First, since the model predicts strong deflation, the monetary authority lowers interest rates relatively rapidly after the technology improvement. In practice, however, inflation may not immediately jump down in response to a technological improvement (for reasons that beg further theoretical explanation), so that monetary policy probably would not ease immediately after the shock. Sluggish inflation appears to be observed empirically (see Fuhrer and Moore, 1995; Roberts, 1998; and Galí and Gertler, 1998). Second, Basu's model does not incorporate many "real rigidities." But Kimball (1995) shows that one can obtain a "contract multiplier" of any desired size by adding real rigidities to the model, thus making price adjustment arbitrarily slow.

improvements might plausibly work through the sticky-price channel. The contractionary effects of sticky prices work through deflation and investment-demand channels, so we first investigate the behavior of those variables. We then study the response of real interest rates and consumption to a technology improvement.

When technology improves, prices should fall, unless the Fed fully accommodates the increased demand for liquidity. This deflation, in turn, tends to be contractionary. Testing this prediction requires that we specify a time-series process for inflation. In the mid-1980s, some commentators argued that the price level is I(2), implying inflation is I(1). By the late 1990s, however, the 1970s look like an aberration—a time when policy responded differently to external shocks. Thus, we model inflation as a stationary process, with a mean shift in the 1970s.[26] We measure the price level with the GDP deflator.

The top right panel of Figure 4 shows response of prices to a technology improvement. (We estimate impulse responses as described in Section IIIB). The price level jumps down on impact, and continues to fall for three years before stabilizing significantly below its initial level. The short-run behavior of prices accords with a model where the Fed does not fully accommodate all shocks. (In fact, the long-run fall in the price level is almost the same as the long-run increase of output, as predicted if the nominal interest rate returns to its pre-shock value and the Fed does not increase the money supply.)

The middle left panel of Figure 4 shows the response of the real interest rate. We measure the real rate as the annual average of the beginning-of-quarter 3-month Treasury bill rate, minus the actual rate of inflation for that quarter. Kimball (1998) shows that in a simple sticky-price model without investment adjustment costs, we should expect *r* to fall initially (due to the fall in investment demand, which reflects that there is excess capital as well as excess labor), then rise as output rises—probably to above its steady-state level—before falling back to that long run level.

The interest rate does tend to fall initially and then rise along with output. When technology improves one percentage point, the real interest rate declines about 0.35 percentage points. The decline is not significant at the 95 percent level, although it is close to significant a year after the shock.

The middle right panel shows the response of consumption. The point estimate shows consumption basically unchanged on impact, before rising strongly in subsequent years. The increase is significant several years following the shock.

---

[26] Results are almost unchanged if we allow a separate mean for all three decades. If we constrain inflation to have a single mean over the period, the estimated short-run behavior of prices also does not change substantially.

This consumption profile can help us distinguish between sticky-price models and flexible-price RBC models with autocorrelated technical change. If one accepts that on impact there is essentially no effect on consumption, that provides further evidence against flexible-price RBC models, even with an arbitrary time-series process for technology. If consumption and leisure are normal, the fall in labor input with unchanged consumption implies that the real wage must fall. But real wages and labor input can both fall only if labor demand shifts back. In a competitive flexible-price model, labor demand depends only on the current level of technology (and the capital stock). Thus, for labor demand to fall, technology must be labor-augmenting (so more technology is like more labor) *and* the elasticity of substitution between capital and labor must be low. For example, with a Leontief production function an improvement in labor-augmenting technology must lead to a short-run fall in labor demand. But the required elasticity of substitution is too low to be credible.[27] In any event, we suspect most RBC theorists would not accept parameterizations implying that technology improvements reduce labor demand.

C. Sectoral Shifts?

Price stickiness can explain why technology improvements are contractionary. Alternatively, even with flexible prices, technology improvements might temporarily reduce output and inputs because of the costs of reallocating resources. The short-run dynamics of this sectoral-shifts explanation depend on the unevenness of technology change across sectors, since inputs may need to shift between sectors in order to find their most profitable use. If this reallocation is costly—as Ramey and Shapiro (1998) document for capital—then technical progress can, in the short run, reduce employment and GDP.[28] In this section, we show that the evidence does not support this alternative hypothesis.

The sectoral shifts hypothesis implies that the larger the dispersion of technology shocks, the greater the pressure to reallocate factors to different uses. Thus, an intuitive test of whether sectoral shifts drive our results is to add measures of the dispersion of technology change to our basic regressions, and see which set of variables has greater explanatory power. A natural dispersion measure, *Disp*, is the cross-sectional

---

[27] With constant returns to scale, labor demand falls only if the elasticity of substitution between capital and labor is less than the output elasticity of capital. Current evidence points to approximately constant returns to scale, so this condition requires an elasticity of substitution of no more than, say, 0.40. Such a small elasticity of substitution contradicts most of the empirical evidence (see, e.g., Pindyck and Rotemberg, 1983).

[28] Lilien (1982) measures reallocative shocks as the cross-industry variance of employment growth, and argues for the importance of sectoral shifts. Abraham and Katz (1986) criticize Lilien's measures. Loungani et al. (1990) and Brainard

standard deviation in technical progress:

$$Disp_t = \left[\sum_{i=1}^{N} w_i \left(\varepsilon_{it} - \varepsilon_t \right)^2 \right]^{1/2}, \tag{5.8}$$

where i indexes industries, ε is the technology impulse as defined in equation (3.2), and w_i is the sector's weight in aggregate value added. Suppose that when technology improves, (measured) input and output fall because of costly factor mobility. It seems reasonable that the greater the dispersion of technology shocks the greater the pressure for reallocation. Thus, we test whether sectoral shifts drive our results by testing whether *Disp* is significant in explaining input and output growth.[29]

It seems unlikely that our technology impulse proxies for effects that are actually due to dispersion, since the correlation between the two variables is quite low. To check this intuition more formally, we run regressions similar to those underlying the impulse-response functions in Figure 3. In particular, in Table 7 we regress output growth and various measures of input growth (dx^V, manhours, and utilization) on estimated technology impulses ε (estimated as residuals from an AR(2) regression of our main technology series from Section III) along with current and two lagged values of *Disp.*

These regressions present econometric issues that also come up with the earlier impulse responses—we generate the regressor $\hat{\varepsilon}_t$ from *dz*, which is itself a generated series. If our regression model (2.2) is correctly specified, the results below are consistent. (Intuitively, generated regressors are like classical measurement error, with the variance of that error tending asymptotically to zero.) However, measurement error in $\hat{\varepsilon}_t$ *does* affect coefficient standard errors, even asymptotically. We do not account for this bias, so the standard errors are not correct.[30] However, these errors are appropriate for testing the joint hypothesis that all the coefficients are zero—if the true coefficient on $\hat{\varepsilon}_t$ is zero, the usual OLS standard errors are correct. In all cases we report, we reject this joint hypothesis at the 5 percent level.

In all cases in Table 7, adding *Disp* has little effect on the coefficients and standard errors of ε and its lags. The timing patterns discussed in Section IIIB are unaltered. Most importantly, the addition of the *Disp*

---

and Cutler (1993) respond to this critique by using the cross-industry dispersion of stock market returns.

[29] This method does not rigorously test the sectoral shifts alternative, since a common aggregate shock affects optimal input use equally in all sectors only if all production and demand functions are homothetic. Nevertheless, even if the dispersion index does not capture all forces leading to input reallocation it should capture some of them. Thus, if sectoral shifts are important, our index dispersion index should significantly predict input and output growth.

[30] Correcting the standard errors in this problem is not straightforward, because the underlying data come from 29

variables leads to only a moderate improvement in the R^2 of the regressions—the increase is between 0.06 and 0.10. Thus, it is unsurprising that we can reject joint significance of *Disp* in all three regressions at conventional significance levels. Overall, the evidence seems more consistent with the sticky-price model of contractionary technology improvement than with the sectoral-shifts alternative.

D. Time-to-Learn?

Several authors have argued recently that technological improvements may reduce growth for a time, as the economy adjusts to new methods of production.[31] For example, Greenwood and Yorukoglu (1997) argue that the introduction of computers caused the post-1974 slowdown in economic growth, since workers and firms had to accumulate new human capital in order to use the new technology effectively. That is, when new technology is introduced, unobserved investment is high; but since the national accounts do not include investments in human capital as output, market output—and hence measured productivity growth—may be relatively low. Therefore, low productivity growth is associated with high input growth, because "full" output is mismeasured. Over time, the investment in knowledge does lead to an increase in measured output and productivity.

This class of models does not generally predict the results we find. Our approach does not correct for the mismeasurement of output caused by unobserved investments in knowledge, so that when technology is introduced, we would conclude (incorrectly) that technology fell. Since measured (as well as unmeasured) inputs are likely to rise at those times, we might find that technology contractions coincide with expansions in inputs. However, with a lag, when market output rises, we would measure a technology improvement—coinciding with a boom. Hence, measured technology improvements would appear *expansionary*, not contractionary. Therefore, if "learning-time" models are important, it should bias our results *against* finding a negative correlation between technology and output and inputs. In addition, Figure 2 suggests that the negative correlation between measured technology and outputs reflects technology improvements as well as declines (relative to trend), so the learning-time story is unlikely to explain our results.

---

individual industries. Hence, "standard" corrections (e.g., Pagan 1984) cannot be applied directly.

[31] See, e.g., Galor and Tsiddon (1997), Greenwood and Yorukoglu (1996), and Greenwood and Jovanovic (1998).

E. "The Cleansing Effect of Recessions?"

Could causality run from recessions to technical improvement, rather than the reverse? For example, if recessions drive inefficient firms out of business, then overall productivity might rise.[32] A difficulty with this hypothesis has been its prediction of countercyclical productivity, while observed productivity is procyclical. One response to this objection is that "other factors (labor hoarding, externalities, etc.) ... make measured productivity procyclical." (Caballero and Hammour, 1994, p.1365). One interpretation of our results is that we have succeeded in controlling for the "other factors," and therefore are finding that technology is countercyclical as the cleansing models predict.

There is a subtle but important point here: With firm-level data, endogenous cleansing would not be a concern. In Basu and Fernald's (1997b) terminology, this effect is a "reallocation"—a shift in resources from inefficient to efficient firms—not a change in firm-level technology. Our theory excludes such effects, since we add up changes in *firm-level* technology to derive aggregate technology *dz*. But since in practice we use industry data, our estimates of sectoral technical change could include *intra*-sectoral reallocation effects, which cleansing models predict are countercyclical.

Even if cyclical reallocations such as cleansing are important, these effects might not affect the cyclicality of our residuals. In particular, to the extent that these kinds of intraindustry reallocations are important, they may be captured in our estimates of industry markups. For example, suppose that for an industry, $dy = \mu dx + R + dz$, where *R* reflects intra-industry reallocations of various sorts. Also suppose these reallocations depend, in part, on input growth *dx:* $R = \delta dx + \xi$. A cleansing effect of recessions implies δ<0; ξ captures any reallocation effects that are uncorrelated with input growth. Even if our instruments are uncorrelated with technology, they may be correlated with reallocations. Suppose ξ is uncorrelated with either the instruments, or any cyclical variables. Then although our estimates of the markup are biased, with a plim of (μ +δ), the estimated technology shocks would not suffer from reverse causation. ξ is then a form of classical measurement error in output, discussed in Section IV. (This cleansing effect could explain why some of our estimated industry markups are less than one.)

However, if ξ is correlated with business-cycle variables—reallocations may, for example, depend on the aggregate cycle as well as sectoral inputs—then some part of our residuals may remain correlated with

---

[32] This idea goes back at least to Schumpeter. Foster, Krizan, and Haltiwanger (1998) provide empirical evidence on

output and input changes for reasons of reverse causality.

The cleansing explanation challenges our basic identifying assumption that industry technical change is exogenous. If we expect at least some of the cleansing effect to work with a lag of more than one year, then Granger causality tests provide some evidence that our results are not being driven by reverse causality. If cleansing is responsible for our results, and some of it operates with a lag of more than a year, we should find that lagged output or input growth significantly predicts our measure of technology change. (It is sensible to expect some lagged effects, since entry and exit of firms is a relatively slow phenomenon.) But we do not find that lagged output or input growth significantly predicts our measures of technology, providing some evidence against the cleansing interpretation.

As we noted above, the cleanest way to distinguish between these hypotheses is to use firm-level data. Most cleansing models take firm-level technical change as exogenous; it is the distribution of inputs across firms with different efficiencies that responds to aggregate demand. Thus, technical change computed using firm-level data are not subject to the cleansing interpretation, and could provide an unambiguous test of our hypothesis. On the other hand, as we note in Section II, there are no firm-level data sets spanning the economy, so a paper using firm-level data could not deal with the aggregate macro issues considered here.

Kiley (1998) points out a second variant of cleansing models, which might be termed models of "recessions as reorganizations"—a term coined by Hall (1991). In these models, firms use times of low demand and output to reorganize production. The reorganization raises productivity at each firm, so even firm-level data do not provide a dispositive test of our hypothesis versus the cleansing alternative. But this variant of cleansing models predicts that when technology improves, investment is also high. The investment may take the form of job search, as in Hall (1991). But we should also observe higher capital investment, as Cooper and Haltiwanger (1996) document for the seasonal cycle in the auto industry.

In a sticky price model, it is easily possible for investment to go down. Indeed, in Kimball (1998), it is declines in investment that cause output to go down in response to a technology shock. This is not surprising, given the relationship of these effects with the behavior of the real interest rate. In a richer sticky-price setting, the behavior or investment is more ambiguous. The short-run ambiguity comes from the fact that firms want less capital now, but know that they will want more in the future. If there are large, convex investment adjustment costs, firms may decide to invest when technology improves in order to spread a given

---

the role of entry and exit in aggregate productivity growth.

volume of investment over a longer period of time. On the other hand, the prices of investment goods will typically be expected to fall as firms reduce prices after the technology improvement—as we showed above, technology improvements are followed by predictable deflations—so firms may choose to wait and buy capital later at a lower price.

Thus, the behavior of investment provides a one-sided test: If investment falls, then the "reorganization" model probably cannot explain our findings.[33] The bottom panels of Figure 4 show the responses of investment and the investment deflator. Investment falls almost four percent in the year that technology improves, significant at the 95 percent level. Investment then recovers strongly, peaking two years after the shock (the peak effect is significant at the 90 percent level). Hence, since investment falls sharply, a flexible-price reorganization model probably cannot explain the results we find. On the other hand, the behavior of the investment deflator is consistent with the predictions of the sticky-price model. Investment goods prices fall about three percent in response to the technology improvement (more than the drop in the overall price level, so the relative price also falls), but the majority of the decline occurs one to two years after the shock. Thus, it is sensible that firms reduce investment after the shock and increase it significantly two years later, when most of the decline in the investment goods price has already taken place.

VI. Conclusion

In this paper, we measure aggregate technology by correcting the aggregate Solow residual for increasing returns, imperfect competition, varying utilization of capital and labor, and aggregation effects. We come to a robust conclusion: in the short run, technology improvements significantly reduce input use while appearing to reduce output slightly as well. Inputs do not recover significantly until about two years after a technology improvement.

These results are inconsistent with standard parameterizations of real-business-cycle models, which imply that technology improvements raise input use at all horizons. We also find that technology shocks do not account for a very high fraction of the variance of inputs and output at cyclical frequencies. By contrast, we argue that these results *are* qualitatively consistent with the predictions of an otherwise-standard dynamic general-equilibrium model with sticky output prices driven by both technology and monetary shocks.

Note that our empirical work actually estimates a composite of the partial effect of a technology improvement and the reactions of policy (especially monetary policy) to that technology shock. If the Fed tries to stabilize inflation, then the size of the effect we estimate should then be regarded as the lower bound on the true partial effect. This point may be especially relevant for estimating the dynamic effects of technology shocks—if the Fed "leans against the wind" and if some part of Fed policy operates with a lag of more than one year, it may appear that the economy recovers more quickly from a technology improvement than would actually be the case without Fed intervention.

We believe that our paper and the identified-VAR literature have identified an important stylized fact: Technical progress is contractionary in the short run, but has its expected expansionary effect in the long run. We advance price stickiness as the major reason for the perverse short-run effect of technical improvement, as does Galí (1999). The evidence is broadly consistent with this view. Nevertheless, it remains possible that other models could be consistent with the evidence as well. Two of the main competing explanations are sectoral-shifts models and "cleansing effects" models. We have presented some evidence that neither is responsible for our findings, but additional, sharper tests are needed before we can be sure that price stickiness is in fact responsible for our results. (A difficulty, of course, is that the alternative hypotheses are not mutually exclusive, but could all contain an element of the truth.) Additional research with firm-level data would be particularly useful.

Nevertheless, if one accepts the view that technology shocks interact with sticky prices, then our results have important implications for monetary policy. First, monetary policy in the United States over the 1950 to 1989 period did not respond sufficiently to technology shocks to allow actual output to adjust quickly to the new level of full employment output. In this light, the debate in the last few years about whether technology has accelerated—and if so, how monetary policy should react—seems very much on target. Short-run movements in technology growth matter just as much for the proper conduct of monetary policy as the long-run rate of technology growth—if not more, since the main concern of monetary policy is short-run stabilization of the economy around the moving target of full employment output. To the extent that looking at a broader information set can give earlier warning about technological movements, monetary policy can be improved in the future.

---

[33] We thank Christopher Foote and Matthew Shapiro for this observation.

References

Abraham, Katharine G. and Lawrence F. Katz (1986). “Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?” *Journal of Political Economy 94, Part 1* (June) 507-522.

Basu, Susanto (1996). “Cyclical Productivity: Increasing Returns or Cyclical Utilization?” *Quarterly Journal of Economics 111* (August): 719-751.

______ (1998). “Technology and Business Cycles: How Well Do Standard Models Explain the Facts?” *Beyond Shocks: What Causes Business Cycles?* Federal Reserve Bank of Boston: Boston.

Basu, Susanto and John Fernald (1997a). “Returns to Scale in U.S. Production: Estimates and Implications.” *Journal of Political Economy 105* (April) 249-283.

Basu, Susanto and John Fernald (1997b). “Aggregate Productivity and Aggregate Technology.” International Finance Discussion Paper No. 593, Board of Governors of the Federal Reserve System.

______ (1999). “Why Is Productivity Procyclical? Why Do We Care?” International Finance Discussion Paper No. 638, Board of Governors of the Federal Reserve System (June).

Basu, Susanto and Miles Kimball (1997). “Cyclical Productivity with Unobserved Input Variation.” NBER Working Paper 5915.

Beaulieu, John J. and Joseph Mattey (1998). “The Workweek of Capital and Capital Utilization in Manufacturing.” *Journal of Productivity Analysis.* 10 (October): 199-223.

Bils, Mark and Jang-Ok Cho (1994). "Cyclical Factor Utilization." *Journal of Monetary Economics 33*, 319-354.

Blanchard, Olivier and Daniel Quah (1989). “The Dynamic Effects of Aggregate Demand and Supply Disturbances.” *American Economic Review* 79 (4): 654-673.

Brainard, S. Lael and David M. Cutler (1993). “Sectoral Shifts and Cyclical Unemployment Reconsidered.” *Quarterly Journal of Economics* 108 (February): 219-243.

Burnside, Craig (1996). “What do Production Function Regressions Tell Us about Increasing Returns to Scale and Externalities?” *Journal of Monetary Economics 37* (April): 177-201.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo (1995). “Capital Utilization and Returns to Scale.” In B. Bernanke and J. Rotemberg, eds., *NBER Macroeconomics Annual 1995* (MIT Press).

______ , _____ , and _____ (1996). “Sectoral Solow Residuals.” *European Economic Review* 40:861-869.

Caballero, Ricardo J. and Mohamad L. Hammour (1994). “The Cleansing Effect of Recessions.” *American Economic Review* 84(December): 1350-68.

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

Cooper, Russell and John Haltiwanger (1996). “Evidence on Macroeconomic Complementarities.” *Review of Economics and Statistics* 78 (February): 78-93.

Domar, Evsey D. (1961). "On the Measurement of Technical Change." *Economic Journal 71* (December): 710-729.

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

Flux, A. W. (1913). "Gleanings from the Census of Productions Report." *Journal of the Royal Statistical Society* 76(6): 557-85.

Foster, Lucia, C.J. Krizan, and John Haltiwanger (1998). "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." NBER Working Paper No. 6803.

Fuhrer, Jeffrey C. and George R. Moore (1995). "Inflation Persistence." *Quarterly Journal of Economics* 110 (February): 127-159.

Galí, Jordi (1998). "Comment." In Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual 1998.* MIT Press: Cambridge, MA.

______ (1999). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89 (March): 249-271.

Galí, Jordi and Mark Gertler (1998). "Inflation Dynamics: A Structural Econometric Analysis." Unpublished, New York University.

Galor, Oded and Daniel Tsiddon (1997). "Technological Progress, Mobility, and Economic Growth." *American Economic Review* 87 (June): 363-82.

Greenwood, Jeremy and Mehmet Yorukoglu (1997). "1974." *Carnegie-Rochester Conference Series on Public Policy*, June, *46:* 49-95.

Greenwood, Jeremy and Boyan Jovanovic (1998). "Solow vs. Solow: Equipment Prices and Growth." Forthcoming in NBER volume *New Directions in Productivity Analysis*, edited by Edwin Dean, Michael Harper and Charles Hulten.

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

Hall, Robert E. (1990). "Invariance Properties of Solow's Productivity Residual." In Peter Diamond (ed.) *Growth,Productivity, Unemployment* (Cambridge, MA: MIT Press).

______ (1991). "Labor Supply, Labor Demand, and Employment Volatility," in Olivier J. Blanchard and Stanley Fischer, eds., *NBER Macro Annual.*

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

Hulten, Charles (1978). "Growth Accounting with Intermediate Inputs." *Review of Economic Studies 45*, 511-518.

Bound, John, David Jaeger, and Regina Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90: 443-450.

Jorgenson, Dale W. and Zvi Griliches (1967). "The Explanation of Productivity Change." *Review of Economic Studies* 34: 249-283.

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

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

Kiley, Michael (1997). "Labor Productivity in U.S. Manufacturing: Does Sectoral Comovement Reflect Technology Shocks?" Federal Reserve Board Manuscript.

______ (1998). "Discussion of 'Are Technology Improvements Contractionary?'" NBER Monetary Economics Program Meeting, April.

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

______ (1998). "The Neomonetarist IS-LM Model." Unpublished, University of Michigan.

______ (1999). "Real Business Cycle Analytics." Unpublished, University of Michigan.

Kimball, Miles S. and Philippe Weil (1997). *Macroeconomics and Finance: A Dynamic and Stochastic Control Approach.* Manuscript.

King, Robert G. and Sergio Rebelo (1998). "Resuscitating Real Business Cycles." Manuscript, University of Virginia.

Lilien, David M. (1982). "Sectoral Shifts and Cyclical Unemployment." *Journal of Political Economy 90* (August) 777-793.

Loungani, Prakash, Mark Rush and William Tave (1990). "Stock Market Dispersion and Unemployment." *Journal of Monetary Economics* 25 (June): 367-88.

Pagan, Adrian (1984). "Econometric Issues in the Analysis of Regressions with Generated Regressors." International Economic Review 25 (February): 221-47.

Pindyck, Robert S. and Julio J. Rotemberg (1983). "Dynamic Factor Demands and the Effects of Energy Price Shocks." *American Economic Review 73* (December): 1066-1079.

Ramey, Valerie A. and Shapiro, Matthew D. (1998). "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie-Rochester Conference Series on Public Policy,* 48 (June): 145-94.

Roberts, John (1998). "Inflation Expectations and the Transmission of Monetary Policy." Manuscript, Federal Reserve Board.

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

Sbordone, Argia (1997). "Interpreting the Procyclical Productivity of Manufacturing Sectors: External Effects or Labor Hoarding?" *Journal of Money, Credit, and Banking 29* (February) 26-45.

Shapiro, Matthew D. (1996). "Macroeconomic Implications of Variation in the Workweek of Capital."

Brookings Papers on Economic Activity (2): 79-119.

Shapiro, Matthew D. and Mark Watson (1988). "Sources of Business Cycle Fluctuations," in Stanley Fischer ed, *NBER Macroeconomics Annual 1988*.

Shea, John (1998). "What Do Technology Shocks Do?" In Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual 1998.* MIT Press: Cambridge, MA.

Solow, Robert M. (1957). "Technological Change and the Aggregate Production Function." *Review of Economics and Statistics 39*: 312-320.

Stigler, George (1939). "Production and Distribution in the Short Run." *Journal of Political Economy 47* (): 305-357.

Taylor, John B. (1993). "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, December 1993, *39,* 195-214.

Tobin, James (1955). "A Dynamic Aggregative Model." *Journal of Political Economy 63* (April) 103-115.

Yorukoglu, Mehmet (1999). "Product vs. Process Innovations and Economic Fluctuations." Paper presented at Carnegie-Rochester Public Policy Conference, April.

Zachariadis, Marios (1999). "R&D-Induced Growth? Evidence from the US Manufacturing Sector." Manuscript, Ohio State University (May).

Table 1. Parameter Estimates

A. Markup Estimates

Durables Manufacturing		Non-Durables Manufacturing		Non-Manufacturing	
Lumber (24)	0.70 (0.10)	Food (20)	0.77 (0.31)	Construction (15-17)	1.18 (0.11)
Furniture (25)	1.05 (0.07)	Tobacco (21)	1.04 (0.28)	Transportation (40-47)	1.61 (0.23)
Stone, Clay & Glass (32)	0.99 (0.08)	Textiles (22)	0.85 (0.08)	Communication (48)	1.55 (0.48)
Primary Metal (33)	0.97 (0.07)	Apparel (23)	1.02 (0.08)	Electric Utilities (491)	1.10 (0.70)
Fabricated Metal (34)	1.17 (0.09)	Paper (26)	1.12 (0.15)	Gas Utilities (492)	0.33 (0.22)
Non-Elect. Machinery (35)	1.36 (0.09)	Printing & Publishing (27)	0.99 (0.22)	Trade (50-59)	0.87 (0.28)
Electrical Machinery (36)	1.04 (0.07)	Chemicals (28)	0.75 (0.15)	FIRE (60-66)	0.04 (0.32)
Motor Vehicles (371)	1.17 (0.04)	Petroleum Products (29)	0.65 (0.14)	Services (70-89)	1.32 (0.31)
Other Trans-port (372-79)	1.01 (0.04)	Rubber & Plastics (30)	1.15 (0.11)		
Instruments (38)	0.87 (0.11)	Leather (31)	0.84 (0.20)		
Miscellaneous Manuf. (39)	1.03 (0.14)				
Column Average	1.05		0.87		0.89

B: Coefficient on Hours Per Worker

Durables Manufacturing	1.25 (0.35)	Non-Durables Manufacturing	0.44 (0.27)	Non-Manufacturing	-0.05 (0.30)

Notes: Coefficients from regression of output growth on input growth and hours-per-worker growth. Estimated by 3SLS for groups (durables, non-durables, and non-manufacturing), with hours-per-worker coefficient constrained to be the same within the group. Instruments are the growth rate of the price of oil deflated by the GDP deflator and the growth rate of real government defense spending (contemporaneous value and one lag), and a monetary shock variable (see text).

Table 2. Descriptive Statistics for Technology Residuals
(Percent per year)

A. Private Economy

	Mean	Standard Deviation	Minimum	Maximum
(1) Solow Residual	1.5	2.0	-3.2	5.7
(2) Tech. Residual (No Util. Correction, Markup Weighted)	1.2	1.6	-3.1	4.8
(3) Tech. Residual (Hours Corrected, Markup Weighted)	1.3	1.4	-1.3	4.2
(4) Tech. Residual (Hours Corrected, Domar Weighted)	2.1	1.3	-0.4	4.6

B. Manufacturing

	Mean	Standard Deviation	Minimum	Maximum
(1) Solow Residual	2.3	3.5	-8.1	8.0
(2) Tech. Residual (No Util. Correction, Markup Weighted)	1.5	3.1	-8.5	7.2
(3) Tech. Residual (Hours Corrected, Markup Weighted)	1.8	2.8	-3.0	8.2
(4) Tech. Residual (Hours Corrected, Domar Weighted)	2.2	2.3	-2.0	7.5

Note: Sample period is 1950-1989. Technology residuals (2), (3), and (4) come from aggregating 29 industry residuals from regressing output on inputs. (2) does not include any controls for unobserved utilization. (3) and (4) include growth in hours per worker to control for unobserved utilization. As described in the text, "markup weighting" weights industry residuals by $w_i / (1-\mu_i s_{Mi})$. Domar weighting weights residuals by $w_i / (1-s_{Mi})$.

Table 3. Correlations of Technology Residuals with Basic Data

A. Private Economy

	Output Growth ***dv*** **(1)**	**Input Growth** dx^V **(2)**	**Hours Growth** ***dh+dn*** **(3)**	**Solow Residual** **(4)**
(1) Solow Residual	0.86***	0.38***	0.49***	1.00
(2) Tech. Residual (No Util. Correction, Markup Weighted)	0.40***	-0.15	-0.08	0.75***
(3) Tech. Residual (Hours Corrected, Markup Weighted)	-0.07	-0.49***	-0.49***	0.32**
(4) Tech. Residual (Hours Corrected, Domar Weighted)	0.18	-0.28*	-0.28*	0.52***

B. Manufacturing

	Output Growth ***dv*** **(1)**	**Input Growth** dx^V **(2)**	**Hours Growth** ***dh+dn*** **(3)**	**Solow Residual** **(4)**
(1) Solow Residual	0.84***	0.37***	0.47***	1.00
(2) Tech. Residual (No Util. Correction, Markup Weighted)	0.42***	-0.14	-0.05	0.79***
(3) Tech. Residual (Hours Corrected, Markup Weighted)	-0.40**	-0.64***	-0.62***	-0.05
(4) Tech. Residual (Hours Corrected, Domar Weighted)	-0.19	-0.49***	-0.47***	0.15

Significance levels: * 10 percent. ** 5 percent. *** 1 percent. (Calculated using Fisher transformation). Notes: See notes to Table 1. Sample period is 1950-1989.

Table 4. Fraction of Variance Due to Technology Shocks

Lags	Output	Inputs	Manhours	Utilization	Solow Res.
0	5	32	36	20	5
1	10	24	29	13	38
3	31	15	21	10	59
10	41	6	14	6	66

Table 5. Correlations Using Electricity-Corrected Residuals

A. Electricity-Corrected Regression Residuals

	Output Growth ***dv***	**Input Growth** dx^V	**Hours Growth** ***dh+dn***	**Standard Produc-tivity**
Tech. Residual (Private Economy)	0.41***	-0.13	-0.08	0.71***
Tech. Residual (Manufact.)	0.23	-0.25*	-0.20	0.60***

B. Electricity-Corrected Solow Residuals

	Output Growth ***dv***	**Input Growth** dx^V	**Hours Growth** ***dh+dn***	**Standard Produc-tivity**
Solow Residual (Private Economy)	0.45***	0.02	0.04	0.64***
Solow Residual (Manufact.)	0.50***	-0.04	0.03	0.82***

Significance levels: * 10 percent. ** 5 percent. *** 1 percent. (Calculated using Fisher transformation).
Sample period is 1950-1989.

Table 6. One-Digit and Industry-Average Correlations

	Corr(*dp*,*dy*)	**Corr(*dp*, *dx*)**	**Corr(*dz*, *dy*)**	**Corr(*dz*, *dx*)**
Construct.	0.38*	0.09	-0.16	-0.44*
Manufact. Durables	0.76*	0.66*	-0.56*	-0.61*
Manufact. Non-Durables	0.55*	0.14	0.58*	0.22
Transport	0.68*	0.15	-0.08	-0.57*
Communications	0.57*	-0.18	0.00	-0.66*
Public Utilities	0.57*	-0.03	0.62*	0.19
Trade	0.79*	0.09	0.82*	0.17
FIRE	0.25	-0.49*	0.47*	0.08
Services	0.82*	0.53*	0.56*	0.19
(Unweighted) Average of One-Digit Correlations	0.59	0.17	0.25	-0.10
Average of 29 Industries (21 Manufact, 8 other)	0.53	0.03	0.33	-0.10

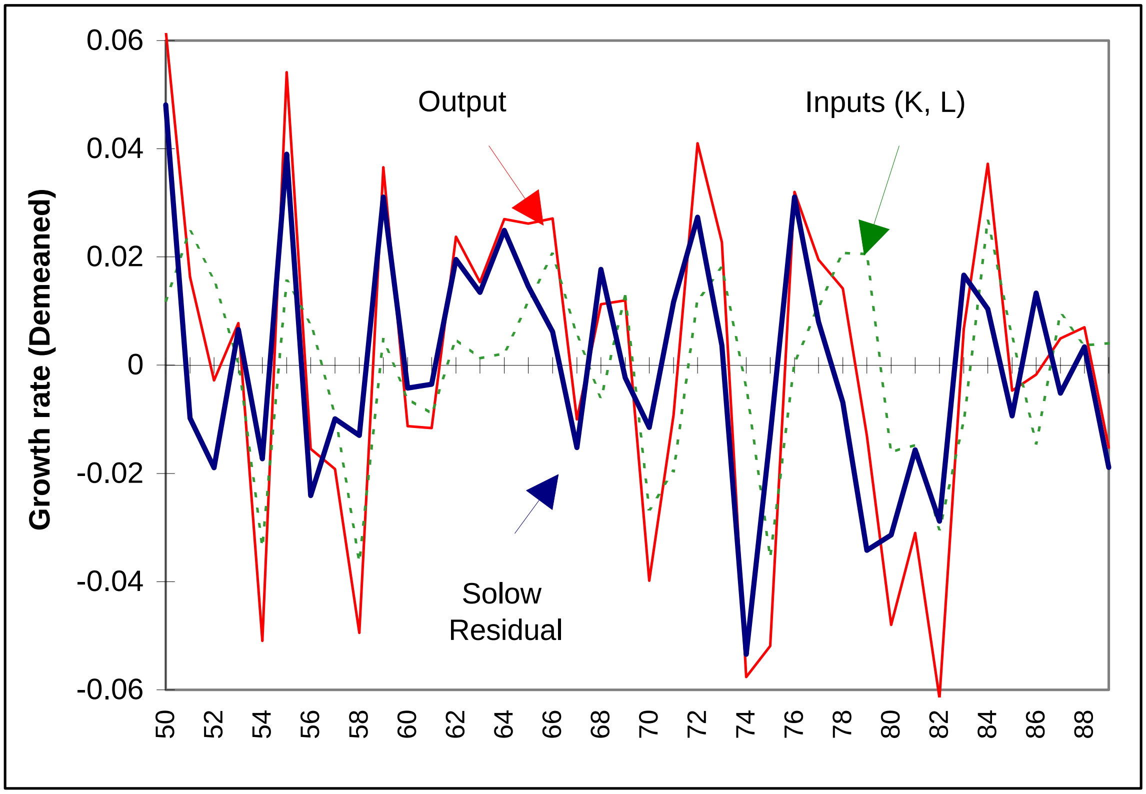
Note: The 29 individual industries and the 9 one-digit industries span the private non-farm, non-mining business economy. Except where noted, all averages are output weighted. Technology *dz* is calculated as the residual from an industry-by-industry gross-output regression as described in Section II, with the full set of corrections for variable utilization. All correlations are calculated from 1950-1989. For one-digit correlations, a * indicates statistical significance at the 95 percent level.

Table 7. Effect of Technology Improvements on Growth Rates of Output, Input, and Utilization

	Output *dv*	**Output** *dv*	**Inputs** dx^V	**Inputs** dx^V	**Hours**	**Hours**	**Utili-zation**	**Utili-zation**
$\hat{\varepsilon}_t$	-0.46 (0.31)	-0.49 (0.32)	-0.68 (0.16)	-0.69 (0.17)	-1.02 (0.23)	-1.01 (0.25)	-0.26 (0.12)	-0.27 0.12)
$\hat{\varepsilon}_{t-1}$	1.17 (0.31)	1.22 (0.34)	0.28 (0.16)	0.24 (0.17)	0.69 (0.23)	0.47 (0.25)	0.45 (0.11)	0.46 (0.11)
$\hat{\varepsilon}_{t-2}$	0.70 (0.31)	0.83 (0.34)	0.45 (0.16)	0.52 (0.17)	0.38 (0.23)	0.73 (0.24)	-0.05 (0.12	0.13 (0.12
$Disp_t$		0.92 (0.64)		0.19 (0.33)		0.42 (0.48)		0.59 (0.23)
$Disp_{t-1}$		0.33 (0.72)		0.44 (0.37)		0.87 (0.53)		-0.07 (0.26)
$Disp_{t-2}$		-0.84 (0.65)		-0.61 (0.34)		-0.91 (0.48)		-0.15 (0.24)
R^2	0.39	0.47	0.47	0.54	0.47	0.57	0.41	0.47
D. W.	2.10	1.95	1.84	1.73	2.08	1.78	1.87	1.82

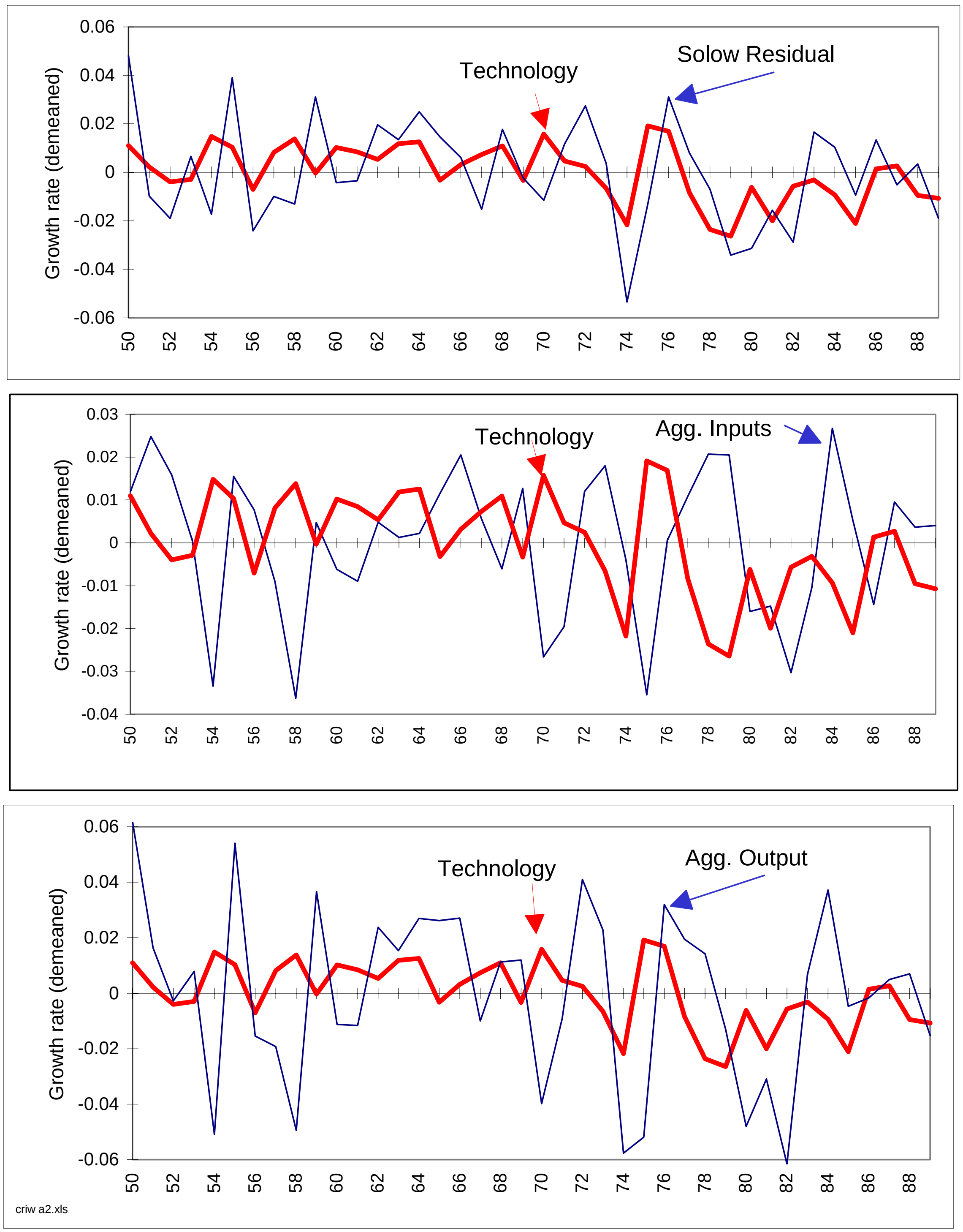
Note: Technology innovations ε_t are residuals from a regression of technology growth (utilization corrected, markup weighted) on a constant and two lags. Regressions in the table include a constant. Sample period is 1954-1989.

Figure 1. Solow Residual, Input Growth, and Output Growth



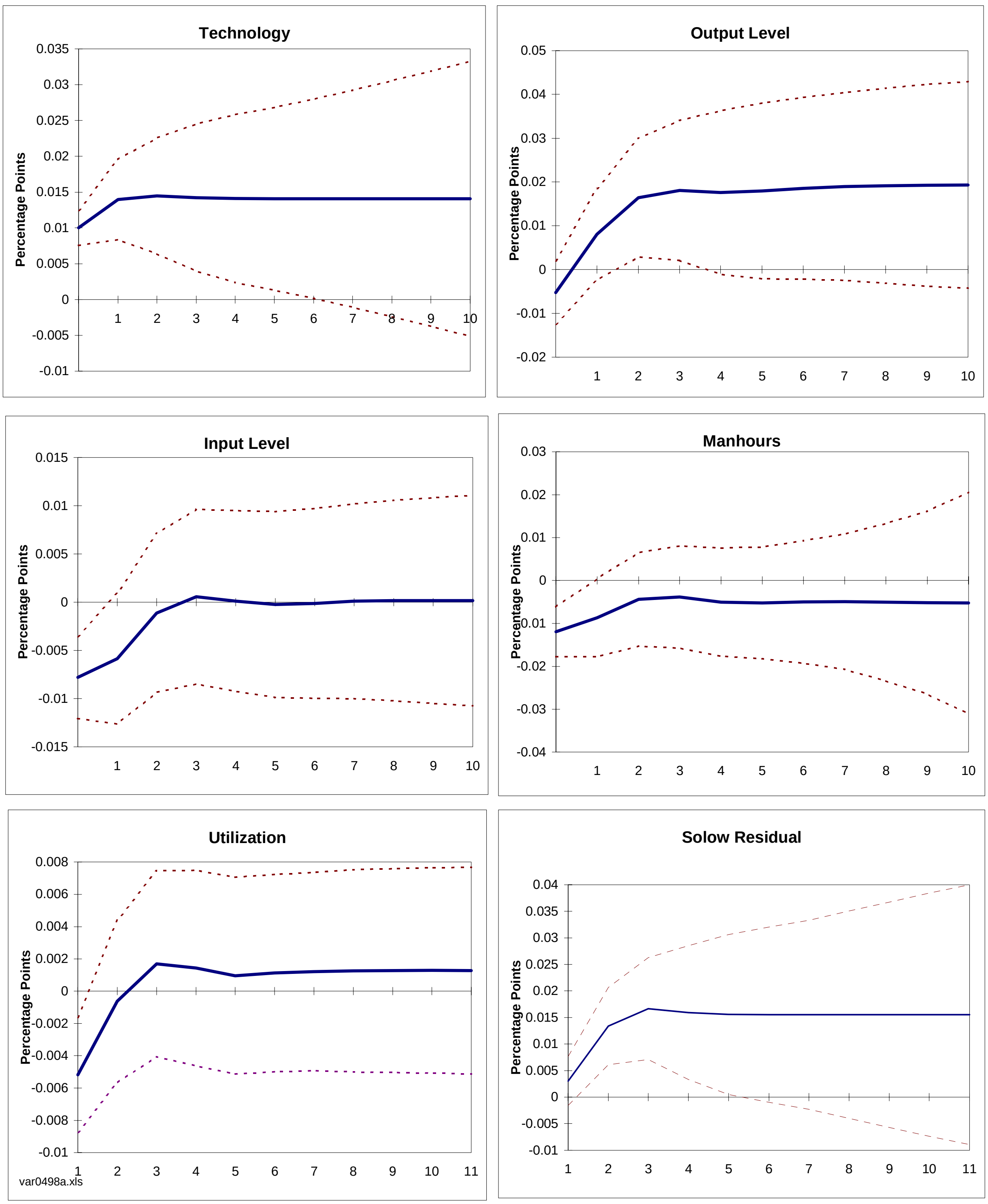
Note: All series are demeaned. Sample period in 1950-89. All series cover the non-farm, non-mining private business economy. Growth in aggregate output is measured as real value added. Growth in inputs is measured as the weighted average of growth in primary inputs of capital and labor. The Solow residual is measured as output growth minus input growth.

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



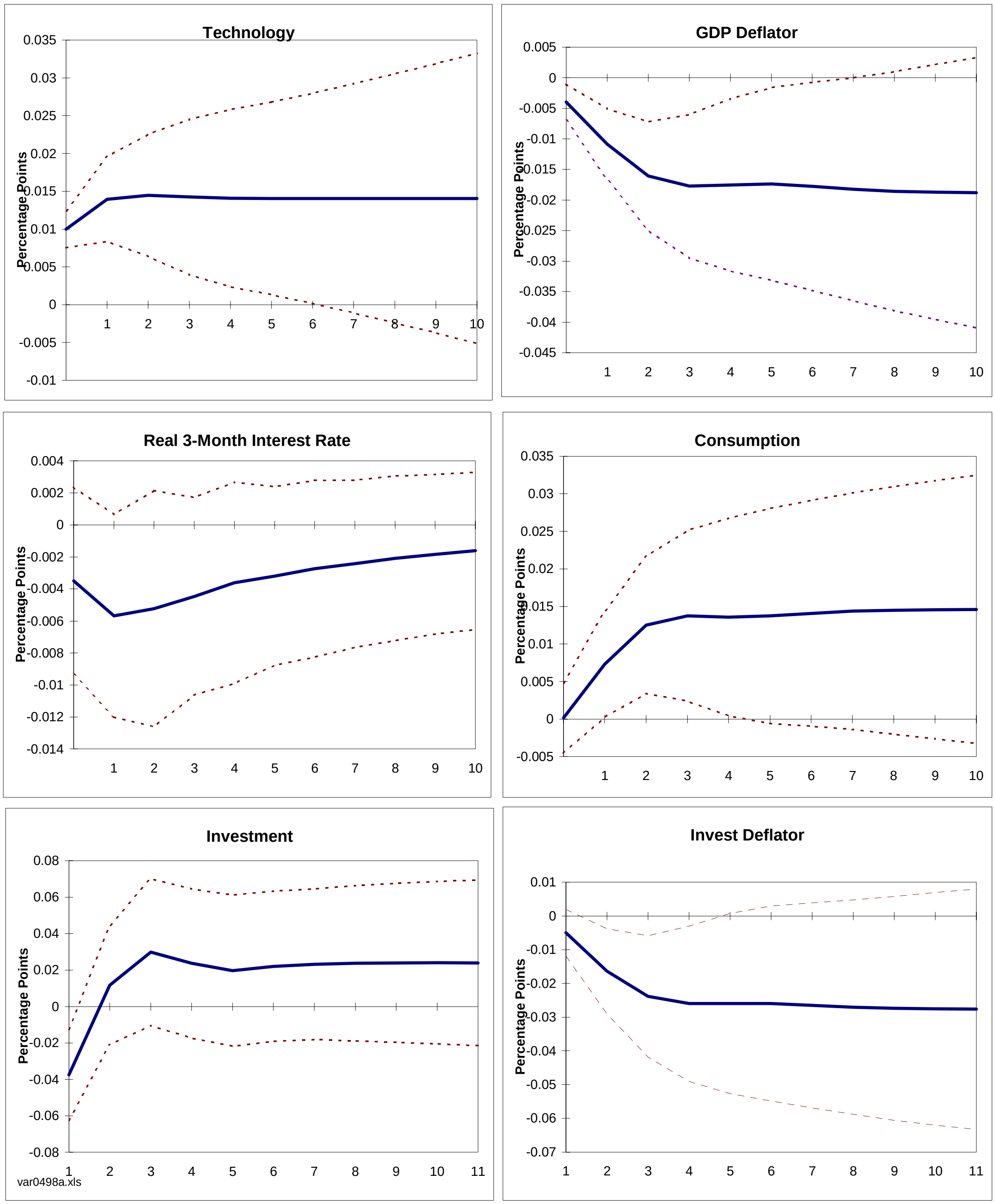
Note: The technology series is the markup-weighted hours-corrected residual. For description of series, see notes to Figure 1. All series are demeaned. Entries are percent changes. Sample period is 1950-89.

Figure 3. Impulse Responses to Technology Improvement: Basic Variables



Note: Impulse responses to a 1 percent (0.01) improvement in technology. Technology is the markup-weighted hours-corrected residual. All entries are percent changes. Dotted lines show 95 percent confidence intervals, computed using Monte Carlo bootstrap method. Sample period is 1952-89.

Figure 4. Impulse Responses to Technology Improvement: Other Variables



Note: Impulse responses to a 1 percent (0.01) improvement in technology. Technology is the markup-weighted hours-corrected residual. All units are percent changes, except for the real interest rate, which is in levels. Dotted lines show 95 percent confidence intervals, computed using Monte Carlo bootstrap method. Sample period is 1952-89