Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 625

September 1998

ARE TECHNOLOGY IMPROVEMENTS CONTRACTIONARY?

Susanto Basu, John Fernald, and Miles Kimball

ARE TECHNOLOGY IMPROVEMENTS CONTRACTIONARY?

Susanto Basu, John Fernald, and Miles Kimball*

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. These results are inconsistent with frictionless dynamic general equilibrium models, which generally predict that technology improvements are expansionary, with inputs and (especially) output rising immediately. However, the results are consistent with plausible 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.

* Fernald is an economist in the International Finance Division of the Federal Reserve Board, and can be reached at Stop 20, Federal Reserve Board, Washington, DC 20551. Basu and Kimball are associate professors of economics at the University of Michigan, and affiliates of the National Bureau of Economic Research; they can be reached at Department of Economics, University of Michigan, 611 Tappan St., Ann Arbor, MI 48109. Email: fernaldj@frb.gov, sbasu@umich.edu, and mkimball@umich.edu. We thank Robert Barsky, Menzie Chinn, Russell Cooper, Dale Henderson, Jordi Gali, Michael Kiley, Lutz Kilian, Robert King, Shinichi Sakata, Matthew Shapiro, and seminar participants at Brown University, Harvard University, the International Monetary Fund, Princeton University, the University of Michigan, the WEA summer meetings, the AEA winter meetings, the NBER Summer Institute, and the NBER Monetary Economics conference. 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.

When technology improves, does employment of capital and labor rise in the short run? Frictionless real-business-cycle models generally predict that it does. By contrast, sticky-price models generally predict that it does *not*: 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-third as much as the Solow residual, a finding that is plausible a priori. 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) the first year, but increases strongly thereafter. Output ultimately increases about as much as technology improves, 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, and our estimates imply that when technology improves, utilization falls sharply on impact before recovering with a lag. That is, utilization falls when technology improves—so the Solow residual rises less than technology. Of course, if technology shocks were the only impulse, we would nevertheless observe a negative (though weakened) correlation between the observed Solow residual and the business cycle. Demand shocks presumably explain why we do not. When demand increases, output and inputs—including utilization—increase as well. We find that shocks other than technology are much more important at cyclical frequencies, so changes in utilization make the observed Solow residual procyclical.

---

[1] For some of the many recent references on technology and the Solow residual, see, for example, Basu (1996), Basu and

We show that these dynamics are quite consistent with the predictions of dynamic general-equilibrium models with sticky prices. Suppose the quantity theory governs the demand for money, so output is proportional to real balances. In the short run, if the supply of money is fixed and prices cannot adjust, then real balances and hence output are also fixed. Now suppose technology improves. Firms now need less labor to produce this unchanged output, so they lay off workers and reduce hours. Over time, however, prices adjust, the underlying real-business-cycle dynamics take over, and output rises. We show in Section V that relaxing the quantity-theory assumption allows richer output dynamics: in particular, output can actually fall after a technology improvement, matching the pattern we observe in U.S. data.[2]

Of course, technology improvements will be contractionary only if the monetary authority does not offset their short-run effects through expansionary monetary policy. This possibility seems particularly plausible in the case of technology improvements, since standard sticky-price models predict that an increase in 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 more likely to be contractionary when the short-run response of inflation is muted, leading the central bank to keep interest rates close to pre-shock levels. This inflation inertia appears to be a fact of the world, though it is not easily generated in sticky-price models.[3] Thus, even with endogenous monetary policy, technology improvements may plausibly generate short-run contractions in output if prices are sticky.

We identify technology shocks using the tools of Basu and Fernald (1997a) and Basu and Kimball

---

Fernald (1997a), Bils and Cho (1994), Burnside (1996), Burnside et al. (1996), and Shapiro (1996).

[2] The economic logic is similar to that of Tobin's insightful (1955) paper, where he argued that:

> "Technological progress has mixed effects. In the absence of monetary expansion and technological progress, price deflation is a necessary concomitant of growth even when the labor supply is increasing just as rapidly as capital. In these circumstances, therefore, growth with stable or increasing employment cannot continue if the money-wage rate is inflexible downward." (p. 113)

Tobin assumed an exogenously-fixed nominal wage. By contrast, the model in Section V has sticky output prices, set by profit-maximizing, imperfectly-competitive firms. In principle, we can thus derive explicit transition dynamics from the short-run, sticky-price equilibrium examined by Tobin to the long-run, flexible-price equilibrium studied in RBC theory. Henderson and McKibbon (1993) study rational-expectations models with nominal wage rigidity in order to examine the optimality of different monetary policy rules in environments with different mixtures of shocks. They argue that the optimal rule changes substantially if technology shocks are important; in particular, their model also predicts that improvements in technology lower employment under some circumstances (but not always). Of course, technology improvements are less likely to have perverse effects in models with sticky wages as opposed to sticky prices: In a sticky-wage model the short-run real wage has to overshoot its frictionless level, which requires prices to fall sharply when technology improves. Thus, it is important to understand the source of nominal frictions in the economy.

[3] Fuhrer and Moore (1995). See also Roberts (1998) and Gali and Gertler (1998).

(1997), who in turn build on Solow (1957) and Hall (1990). Basu and Kimball allow for variable capital and labor utilization, increasing returns to scale, and markups of price over marginal cost. Their basic insight is that since cost-minimizing firms operate simultaneously on observed and unobserved margins, observed margins can proxy for unobserved changes in utilization. For example, if labor is particularly valuable, then firms work existing employees both longer (increasing observed hours per worker) and harder (increasing unobserved effort). Basu and Fernald, by contrast, emphasize sectoral heterogeneity. They argue that under a variety of conditions—for example, different industries having different degrees of market power—inputs may have different marginal products in different uses. Aggregate productivity growth then depends on which sectors change input use the most over the business cycle.

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.

Gali (1998) 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, Gali and Kiley make the identifying assumption that only technology shocks can affect labor productivity in the long run. Gali 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, measuring technology innovations by changes in R&D spending and patent activity. He splits his sample by product and process innovations, since the latter are the only type that can reliably be detected in output and TFP data. He finds that process innovations increase TFP with a lag of several years. As TFP rises, inputs fall significantly—the result we find.

Relative to these papers, we make three contributions. First, we refine the theoretical arguments about how technology shocks affect inputs and output, showing that in principle, output may fall when technology improves. Second, our results do not depend on long-run identifying assumptions that may not hold. For example, Gali'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.[4] 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. Third, 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 the manufacturing sector.

Despite these issues, the three approaches are best regarded as complements, with distinct identification schemes and strengths. Despite differing data, countries, and methods, we find similar results. It thus appears we have uncovered a robust stylized fact.[5] Our findings support the predictions of sticky-price business cycle models but are not consistent with the usual predictions of frictionless RBC models, and hence have strong implications for business-cycle theory.

The paper has the following structure. Section I reviews our method for identifying sectoral and aggregate technology change, largely following Basu and Kimball (1997). Section II discusses data and econometric method. Section III presents our empirical results. Section IV discusses robustness checks. Section V presents different interpretations of our results, including our preferred sticky-price interpretation, which extends Kimball's (1995) model. 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 our method of aggregation, while Subsection B discusses the theory underlying our method of controlling for utilization.

A. Firm and Aggregate Technology

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

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

[5] In private communication, Jordi Gali informs us that in U. S. data his VAR-based measure of technological change has a statistically significant correlation of 0.6 with our technology residual.

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

The firm produces gross output, Y_i, using the stock of capital K_i, labor employees N_i, and intermediate inputs of energy and materials M_i. We assume that the capital stock and the number of employees are quasi-fixed, so that their levels cannot be changed costlessly. However, firms may vary the intensity with which they use these quasi-fixed inputs: A_i represents variations in the utilization of capital; H_i represents hours worked per employee; and E_i represents the effort of each worker at the firm. Total labor input, L_i, is the product $E_iH_iN_i$.[6] 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.

Hall (1990) considers the cost-minimizing behavior of an imperfectly competitive firm with no variations in intensity of input use. Basu and Kimball (1997) extend Hall's approach to the case of variable input utilization. Conceptually, they separate total input growth into the sum of observed revenue-share-weighted input growth, dx_i, and unobserved growth in utilization, du_i. (For any variable J, we define dj as its growth rate.) If the firm is imperfectly competitive, it charges a markup μ_i of price over marginal cost. Basu and Kimball essentially write output growth, dy_i, in terms of total inputs, the markup, and gross-output-augmenting technology change, dz_i. In particular, their derivations imply that

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

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 share of payments to input J in 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

---

[6] Our empirical procedure does not constrain the exponent on H to equal one, as Basu and Kimball (1997) discuss.

micro definition is an appropriately-weighted average of firm-level technology change. With constant returns and perfect competition, these two perspectives are equivalent (Domar, 1961; Hulten, 1978). Rotemberg and Woodford (1995) show that equivalence also holds with imperfectly competitive product markets, 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; that is, aggregate technology from a macro perspective may not be a weighted average of firm-level technology. 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 is the firm's share of aggregate nominal value added:

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

Conceptually, this measure first converts the gross-output technology shocks to a value-added basis by dividing through by $1-\mu s_M$. (A value-added basis is desirable because of the national accounts identity, which tells us that aggregate final expenditure equals aggregate value added.[7]) 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. Results are qualitatively unaffected (and quantitatively barely changed) by using one measure rather than the other.

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:

---

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

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

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 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. To do so, Basu and Kimball (1997) use the basic insight that a cost-minimizing firm equates the marginal benefits of all factors of production to their marginal costs, regardless of whether these inputs are observed or unobserved. Thus, increases in observed inputs can, in principle, proxy for unobserved changes in utilization. This insight allows them to control for variable utilization without 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. Their results use only the cost-minimization problem and the assumption that firms are price-takers in factor markets; they do not require any assumptions about the firm's pricing and output behavior in the goods market.

Assuming that capital utilization (the number of shifts) is a continuous variable, Basu and Kimball work with the following problem:

$$\underset{E,A,H,I,R,M}{\text{Min}} \int_0^\infty e^{-\int_0^t r\, d\tau}\left[WLG(H,E)V(A) + P_M M + WN\Psi(R/N) + P_I I J(I/K)\right] dt \qquad (1.7)$$

subject to

$$\bar{Y} = F(AK, EHN, M; Z) = Z\Gamma\left((AK)^{\alpha_K}(EHN)^{\alpha_L} M^{\alpha_M}\right) \qquad (1.8)$$

$$\dot{K} = I - \delta(A)K \qquad (1.9)$$

$$\dot{L} = R. \qquad (1.10)$$

I is gross investment; R is hiring net of separations; $V(A)$ is the shift premium; $WG(H,E)V(A)$ is total compensation per worker as specified by an implicit contract; $WN\Psi(R/N)$ is the total cost of changing the number of employees; $P_I I J(I/K)$ is the total cost of investment; and $P_M M$ is the cost of materials.

F is a generalized Cobb-Douglas production function; Z is the gross-output-augmenting technology level; $\delta(U)$ is the variable rate of depreciation; and α_K, α_L, and α_M are the shares of the inputs in total cost.

Basu and Kimball assume that if workers must work longer hours or put in more effort per hour worked, firms must compensate them for their disutility with a higher wage; hence, roughly speaking, the function G is assumed to be convex in both effort and hours worked.[8] This wage is an implicit contract over time and states of nature, and is not observed at high frequencies. The setup allows for two possible costs to increasing capital utilization: wear and tear, and a shift premium. That is, capital may depreciate in use, and workers may be paid more to work the night-shift (or, in general, at any less-preferred times). Both costs are assumed to be convex: thus they assume $\delta' > 0$, $\delta'' > 0$ and $V' > 0$, $V'' > 0$.

Under these assumptions, Basu and Kimball relate utilization growth to three observed variables:

$$\mu du = a dh + b(dp_M + dm - dp_I - dk) + c(di - dk). \qquad (1.11)$$

dp_M is the growth in the price of materials, dp_I is the growth in the price of investment goods, and di is the growth in investment. The coefficients a, b, and c are complex combinations of underlying structural parameters. Basu and Kimball derive equation (1.11) in their Appendix A; here, we explain the intuition for these variables, by considering two polar assumptions regarding the cost of higher capital utilization.

As one polar case, assume that there is no shift premium. The hours per worker term in equation (1.11) then controls only for unobserved variations in effort. Suppose a firm finds that the marginal revenue product of labor exceeds the wage. This firm would like to hire more workers, but with quasi-fixed labor it cannot adjust the stock of workers, N, immediately. It has two choices: work current employees longer, or harder (compensating appropriately in both cases). A cost-minimizing firm will choose to do both, so that at the margin it costs the same to increase labor input along either margin. Hence, when observed hours per worker are high, unobserved effort should also be high.

The second and third terms in equation (1.11) reflect variations in capital utilization coming from "wear and tear," that is, depreciation of capital in use. The second term reflects the benefit of higher capital utilization. Intuitively, capital's marginal product is high when it is scarce relative to other inputs, as reflected in this term by the relative value of materials and the stock of capital. That is, higher capital

---

[8] 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

utilization allows the firm to save labor and materials; equation (1.11) uses the first-order condition between utilization and materials input, since materials are probably better measured than total labor input. The third term reflects the cost of using up valuable capital through faster depreciation. For a given depreciation function, this cost is proportional to the current shadow value of capital, marginal Q. Marginal Q is not observed directly, but a first-order condition equates marginal Q to the investment-to-capital ratio. Hence, the expected sign on the third term is negative: high I/K signals a high shadow value of installed capital, and hence, a high cost of the cost of wear and tear from higher utilization.

As a second polar case, assume that the sole cost of higher capital utilization is a shift premium, so that capital does not depreciate in use. Hours per worker then proxies for capital utilization as well as labor effort, since shift premia create a link between capital hours and labor compensation. The 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. In this plausible case, the only additional variable needed to control for changes in utilization of capital and labor is the change in hours per worker. We use this parsimonious case to derive one of our series of utilization-adjusted residuals.

Finally, suppose that capital utilization is costly to the firm both because it increases the marginal rate of depreciation and because it raises the shift premium. In this case, Basu and Kimball establish that the three variables in equation (1.11) continue to proxy for all changes in utilization. This result is intuitively reasonable, since this case is a convex combination of two cases in which these three variables or a subset of them are sufficient proxies for variations in utilization.

So far we have discussed only model-based proxies, using cost-minimizing conditions derived under fairly general assumptions. One strand of the literature controls for variations in utilization by assuming fixed proportions between an observed input and unobserved inputs. 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:

$$du + dk = d\left(electricity\right). \tag{1.12}$$

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.

---

guarantees a global optimum; assuming $\Phi_{11} > \Phi_{12}$ and $\Phi_{22} > \Phi_{12}$ ensures that optimal H and E move together.

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 U.S. economy, and our "firms" are 33 industries; for manufacturing, these industries correspond roughly to the two-digit SIC level.

Each industry contains thousands or tens of thousands of firms, so it may seem odd to take industries as firms. Unfortunately, no firm-level data sets span the economy. In principle, 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. These data consist of a panel of 33 private industries (including 21 manufacturing industries) that cover the entire U. S. non-farm 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).) These data are available 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, distinguished by the BEA in constructing the national product accounts. Hence, for each of these 53 assets, indexed by s, the user cost of capital is

$$R_S = \left(r + \delta_S\right)\frac{\left(1 - ITC_S - \tau d_S\right)}{\left(1 - \tau\right)}, \text{ s} = 1 \text{ to } 53. \tag{2.1}$$

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.[9] (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 "policy shocks" from an identified VAR. We sum the four quarterly policy shocks in year *t-1* as instruments for year *t*.[10]

B. Estimating Technology Change

To estimate "firm-level" technology change, we combine equations (1.2) and (1.11) for each industry. Although we could estimate these equations separately for each industry (and indeed do so as a check on results), some parameters—particularly on the utilization proxies—are then estimated rather imprecisely. To mitigate this problem, we combine industries into four groups, estimating equations that restrict the utilization parameters to be constant within industry groups. Thus, for each group we have

$$dy_i = c_i + \mu_i dx_i + a dh_i + b(dp_{Mi} + dm_i - dp_{Ii} - dk_i) + c(di_i - dk_i) + dz_i. \qquad (2.2)$$

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); mining and petroleum extraction (4); and all others, mainly services and utilities (8).

---

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

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

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 above.

After estimating equation (2.2), the sum of the industry-specific constant c_i and residual dz_i measures technology change in the gross-output production function. Since we are ultimately interested in the aggregate effects of technology shocks, we take an appropriately weighted average of the firm-level estimates of technology change, using equation (1.4).

We also estimate two special cases of equation (2.2) at an industry level. First, we drop all utilization terms (i.e., setting $a = b = c = 0$). Once we aggregate residuals across industries, we obtain a non-utilization-adjusted series for technical change, which corrects only for the aggregation effects emphasized by Basu and Fernald (1997a,b). Second, we estimate (2.2) using only the hours correction:

$$dy_i = \mu_i dx_i + adh_i + dz_i \qquad (2.3)$$

This parsimonious equation allows us to control for both capital and labor utilization if the sole cost of higher capital utilization is a shift premium.

As a robustness check, we also use electricity use as a proxy for du. This approach ignores variations in labor effort de, and amounts to combining equations (1.2) and (1.12).

Finally, we confirm that our results are robust to using industry-by-industry rather than group estimation. Although the variance of the estimated technology residuals increases significantly, the main correlation results in the next section are not qualitatively affected.[11]

III. Results

A. Basic Correlations

Table 1 reports summary statistics for four series: (i) the Solow residual; (ii) a series that makes no utilization corrections, but corrects only for aggregation biases; (iii) a "technology" measure based on

---

[11] 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 fully corrected Domar-weighted residuals turn out to have a correlation of 0.96 with the fully-corrected residuals below, with similar variances and correlations with inputs and outputs.

(2.3), which uses the hours-per-worker correction only; and (iv) a "fully corrected" technology measure, based on equation (2.2). Note that the first measure uses aggregate data alone, whereas the other three are based on sectoral regression residuals, which are aggregated using equation (1.4).

Panel A shows results for the entire economy. Our corrected series have about the same mean as the Solow residual. However, the variance is substantially smaller: The variance of the fully-corrected series is less than one-third that of the Solow residual, so the standard deviation (shown in the second column) is only about 55 percent as large. The reported minimums show that we do estimate negative technical change in some periods, but the lower variance of the technology series implies that the probability of negative estimates is much lower. For example, the Solow residual is negative in 12 out of 40 years; the fully-corrected residual is negative in only 5 out of 40 years.

Panel B gives results within manufacturing alone. Data within manufacturing (especially for output) are often considered 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 fully-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, thereby reducing the Solow residual relative to the technology series. The phase shift, in particular, appears to reflect primarily movements in hours per worker, which generally increase a year after a technology improvement. In the Basu-Kimball model, increases in hours per worker imply increases in unobserved effort, which in turn increase 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 present a sticky-price model 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 impact of technology on aggregate variables in a more formal manner. We first study simple correlations between variables. To our original three business-cycle variables we add the growth of total hours worked ($dh + dn$).

Table 2 shows the usual business-cycle facts for these four variables: output and inputs are strongly positively correlated, and all are positively correlated with the Solow residual. 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 95 percent confidence intervals show that all are significant.

Table 3 contains the key results of the paper, the correlations between our technology measures and business-cycle variables. Panel A shows results for the aggregate private economy. The correlations with aggregate technology change differ sharply from those predicted by the usual RBC model (e.g., Cooley and Prescott, 1995). With full corrections, 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 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.

The correlations for the hours-adjusted series are even more strongly negative than those of the fully-adjusted series. This result may reflect the procyclicality of investment. In particular, if capital depreciates in use, then capital utilization should be less variable than implied by the hours correction alone. That is, high investment in booms indicates a high value of installed capital (not merely a high current marginal product), so it is costly to wear it out through higher utilization. Thus, the fully-adjusted series actually makes a smaller utilization correction than the hours-corrected series.

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.[12]

Panel B of Table 3 presents similar results for manufacturing alone. The two utilization-corrected series show large negative correlations between technology change and both output and input growth. Compared with the results in Table 3A, the correlations for the utilization-corrected series with output are much more negative and with the Solow residual are much closer to zero. The input correlations are generally large and negative; all are statistically significantly different from zero at the 95 percent level.

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 find the innovations, ε:

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

For *dz* we use our fully-corrected measure of technology change. To derive the impulse response of any variable *J* to a technology 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*. We estimate the system by SUR, then use it to compute the impulse responses. In all cases we use a lag length of two periods (in our case, years).[13]

---

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

[13] 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

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.[14]

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 (manhours, utilization, or dx^V). The fall in output is not statistically significant.

Output grows strongly after the shock: two years out, the impulse response differs significantly from zero, with output rising about 1.8 percent. Inputs grow more slowly, but the standard errors of the estimates are large. For example, the point estimates say that dx^V falls 0.8 percent on impact, and then recovers to its pre-shock level (normalized to zero) in three years. However, at three years the 95 percent confidence interval runs from about 1 percent to -1 percent. The same is true of manhours, although the point estimate of manhours never recovers to its preshock value. The point estimates show utilization remaining above its pre-shock level indefinitely.

That technology improvements reduce both output and inputs on impact contradicts standard flexible-price RBC models driven by permanent technology shocks.[15] Of course, if technology-growth is

---

permanent changes in the size of the labor force that are not related to technology.

[14] 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*.

[15] Kimball (1998) proves that a permanent improvement in labor-augmenting technology cannot lead to a fall in both output and labor input in a basic RBC model. (A "basic RBC model" has aggregate output given by the constant returns to scale function $Y = ZL\,f\left(K/ZL\right)$, $f' > 0, f'' < 0$; time discount rate ρ; felicity given by $u\left(C, L\right)$, where C is consumption and u a function that yields normality of both consumption and leisure (inferiority of labor); and capital accumulation obeying $\dot{K} = Y - C - \delta K$.) A sketch of the proof follows. (1) *In the new steady state after the technology improvement, the level of output (both net and gross) is higher than in the original steady state.* (a)

strongly positively autocorrelated, 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. We return to the issue of autocorrelated technology change when we discuss consumption in Section V.

In a standard RBC model (e.g., Cooley and Prescott, 1995) with a capital share of 0.35, a 1.4 percent increase in Hicks-neutral technology should increase output by about 2.2 percent in the long run (computed as 1.4/(1 - 0.35)), inputs (including capital) by about 0.8 percent in the long run, and leave manhours and utilization unchanged. 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.

The short-run effects of technology improvements contrast sharply with 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

---

Since $f'\left(K/ZL\right) - \delta = \rho$ in steady state, the new steady state has the same ratio *K/(ZL)* and a real wage *Z* times as high as in the old steady state. (b) With the net marginal product of capital positive and *K* and *ZL* moving in tandem, steady state *K, ZL,* output *Y* and net output $Y - \delta K$ must move together. (c) If *K*, *ZL*, and net output $Y - \delta K$ fell, with *Z* higher, steady state *L* would have to be lower. (d) Normality of both consumption and leisure implies that for a given real wage, a fall in *L* is associated with a rise in *C*; *a fortiori,* with the real wage *W* rising, a fall in *L* must be associated with an increase in *C*. (d) If net output $Y - \delta K$ fell while *C* rose in a closed economy with no change in government purchases, net investment would have to fall—but by definition, net investment is zero at the steady state. Therefore, *K*, *ZL*, *Y* and $Y - \delta K$, which must move together, cannot each be lower in the new steady state and must each (at least weakly) increase. (2*) Immediately after the impact of the permanent technology improvement, output cannot fall.* (a) If output were to fall, since *K* is unchanged at that point, *ZL* would have to fall, and the real wage would have to rise while *L* fell. (b) By normality of consumption and leisure, a fall in *L* coupled with a rise in the real wage must be associated with a rise in consumption. (c) With no change in δK, a fall in output combined with a rise in consumption implies a fall in net investment. (d) A fall in net investment in response to the permanent technology shock is inconsistent with being on the saddle path to a new, higher steady-state level of capital. Therefore, output must rise immediately following a permanent technology improvement.

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 ensuring that the model impulse responses match those that we and others find.

Tables 5 to 7 explore the robustness of the regression results underlying Figure 3, confirming the robustness of our basic conclusions to the three different measures of technology change. We regress business-cycle variables on the contemporaneous technology innovation, $\hat{\varepsilon}_t$, and two lags. (Coefficients on longer lags are generally small and insignificant.) Since we assume technology change is exogenous, we use OLS regressions. The coefficients are elasticities.

These regressions present econometric issues that also come up with the 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.[16] 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.

Table 5 relates technology change to primary-input use, dx^V. In all cases, technology improvements significantly lower input use initially. The uncorrected technology measure shows inputs rising significantly one year after a technology improvement, but the fully-corrected and hours-corrected measures show that the increase is delayed another year. The R^2 generally exceeds 0.4, implying that technology shocks substantially affect the variance of inputs. However, the contemporaneous effect of a technology improvement is the opposite of what standard RBC models predict.

Compared with the other two series, why does the non-utilization-adjusted series show a larger

---

[16] Correcting the standard errors in this problem is not straightforward, because the underlying data come from 33

positive effect of a technology improvement on input use with a one-year lag? The unadjusted series series does not subtract changes in utilization, so that if utilization and observed inputs increase together in response to a lagged improvement in true technology, it will appear that $\hat{\varepsilon}_{t-1}$ has a larger effect on observed inputs than is actually the case. We confirm this statement in Table 6, where we regress changes in utilization on technology. We use two measures of utilization change: the fully-corrected measure, which allows for depreciation in use, and the hours-corrected measure, which does not. Aggregate utilization change is a weighted sum of firm-level utilization change, as in equation (1.6).

Table 6 shows that utilization—a form of primary input—responds qualitatively the same way as observable primary inputs. That is, when primary input use falls in response to a technology improvement, capital and labor utilization also fall. That is, Table 6 suggests that on impact, technology improvements lead to a significant fall in utilization; with a lag of up to two years, technology improvements are associated with increases in utilization. As suggested by the earlier results in Table 5, utilization increases most in response to a one-year lagged technology shock.

The results in Table 6 also 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. 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.

Table 7 presents similar regressions relating technology change and output growth. Here, the phase-shift shows up as a timing difference between the results using the utilization-corrected and uncorrected residuals. The corrected series show output falling in the first year, although the effect is never statistically significant. The corrected series then show output growing strongly in the two subsequent years. On the other hand, the uncorrected series shows output rising strongly the first two years. However, since the uncorrected series does not control for utilization change, it shows maximum "technology improvement" in the year *after* technology changes, when utilization increases sharply. The uncorrected series misinterprets this increase in utilization as an improvement in technology.

---

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

IV. Robustness checks

We now check the robustness of our basic results to the method of controlling for utilization, and to classical measurement error. We also study the properties of technology shocks at the sectoral level.

A. 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 8 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. Following Burnside et al., output is value added and inputs are a weighted average of capital and manhours. 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 2, it is close to zero or negative. Only in manufacturing, with regression residuals, can we not reject the hypothesis of a zero correlation. However, 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.

Our results differ quantitatively from those of Burnside et al. (1996), who find that the electricity correction reduces technology's correlation with output to 0.1 in manufacturing. Differences in 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. By contrast, the hours-per-worker proxy appears quite suitable for such industries. Our results suggest that omitting this effort correction changes results noticeably. However, for the manufacturing sector alone, results based on electricity-corrected residuals agree fairly closely with ours.

B. Classical Measurement Error

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. We explore this issue with a very simple model that abstracts from variations in utilization and does not explicitly consider aggregation across industries; neither changes the basic message that measurement error probably cannot explain our results.

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:

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

and

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

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:

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

and

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

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 of a negative correlation between output growth and technology change in 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.

However, 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 for exposition, our technology change series takes a weighted average of technology shocks across sectors. To the extent that measurement error is independent across industries, using 33 industries considerably attenuates the various biases.

C. 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 9 present results for 10 (approximately one-digit) industries, as well as average correlations for the 33 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 2. 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 repeat the specification from Table 5—estimated as a system of seemingly unrelated regressions, with coefficients on current and lagged sectoral technology innovations constrained to be the same across industries—and again 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 Gali's (1998) identification scheme.[17] All but three of the 20 correlations are positive; seven of the positive correlations are significant at the five percent level.

V. Interpretations of the Results

A. Price Stickiness

In calibrated real-business-cycle models with flexible prices, technology improvements generally increase both both inputs and output immediately. These models explore whether technology shocks lead to comovement that matches the stylized facts of business cycles. Given that the central stylized fact of business cycles is the comovement between inputs and output, if technology shocks drive the cycle then almost any sensible calibration implies that technology improvements increase inputs and output.

In a sticky-price model, by contrast, technology improvements can easily be contractionary. Consider the easiest case, where the quantity theory governs the demand for money and the supply of money is fixed. If prices are sticky in the short run, then real balances are also fixed in the short run. Now suppose technology improves. Since the price level is sticky and demand depends on real balances, output does not change in the short run; with the improved technology, firms need fewer workers to

---

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

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 might cause output itself 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 inflation (since the price level must fall over time), which in turn increases money demand (for a given real interest rate). Output then must fall to restore equilibrium in the money market.

The rest of this section explores this argument in more detail, focusing on reasons why improvements in technology have an ambiguous effect on output and, to a lesser extent, on factor inputs. These aspects are common to virtually any dynamic general-equilibrium model with sticky prices. Our discussion here is theoretical, but Basu (1998) shows that a calibrated DGE model with staggered price setting reproduces quite accurately the impulse response to technology shocks that we find in the data.

Our presentation builds on Kimball's (1995) Neomonetarist model, which adds imperfect price flexibility and imperfect competition to an RBC model. We focus on equilibrium conditions in the markets for money and for renting capital. These markets do not summarize the whole model, but understanding equilibrium in these two markets sheds much light on the properties of the full model.

A simple graph, showing short-run equilibrium in the Kimball (1995) model, can clarify the theoretical issues that affect whether technological improvements are contractionary in the short-run. Consider the benchmark case with no investment adjustment costs. Figure 4 shows the intersection of the net rental rate (NRR) curve, giving the rental rate of capital (net of depreciation) as a function of the level of output and other variables, and the monetary policy (MP) curve, giving the real interest rate as a function of the level of output, the level of inflation, and other variables.

We first discuss the NRR curve. By arbitrage between physical and financial investment, the absence of investment adjustment costs requires that the rental rate of capital net of depreciation, $R-\delta$, equal the real interest rate r whenever gross investment is positive—as we assume here. With imperfect price flexibility and imperfect competition, one must derive the rental rate for the capital services from cost-minimization rather than from profit maximization. Indeed, since profit maximization in any ordinary sense requires optimal price setting, imperfect price flexibility implies that firms may not maximize profits in the short run. When minimizing costs for a given output level, the value of additional capital services reflects the value of variable inputs that are saved if inputs of capital services rise. It suffices for exposition to assume a Cobb-Douglas production function with capital and labor, but results generalize easily to more general homothetic production functions. We allow labor to stand for all variable factors, including such control variables as effort per hour and the number of shifts; we needed the complications introduced earlier in the paper for the empirical task of identifying true technology shocks from the data, not for the theoretical point about the effect of technology shocks on models with imperfect price flexibility. Let α_K be capital's share in costs and α_L be labor's share in costs. Then by definition

$$\frac{\alpha_K}{\alpha_L} = \frac{RK}{WL}$$

where W is the real (shadow) wage and L is the quantity of labor. Solving for the rental rate of capital services net of depreciation (which should equal the real interest rate r in short-run equilibrium),

$$r = R - \delta = \frac{\alpha_K}{\alpha_L}\frac{WL}{K} - \delta. \tag{5.1}$$

This NRR curve slopes upward, since for given technology and capital, producing more output requires more labor; this requires a higher real wage W to get the higher quantity of labor L, as the fully-optimizing representative household moves up along its Frisch labor supply curve.

What happens to the NRR curve when technology improves? The marginal utility of consumption, λ, determines the location of the Frisch labor supply curve as a summary statistic for wealth and interest rate effects. Integrating the standard Euler equation $(d\lambda/dt)/\lambda = \rho - r$, where ρ is the utility discount rate, yields the useful equation

$$\ln \lambda_t = \ln \lambda_\infty + \int_t^\infty \left(r_\tau - \rho\right) d\tau. \tag{5.2}$$

As in Kimball (1995), let us make the useful analytical approximation that prices adjust fast enough that we can neglect the effects of the short-run movements in the real interest rate on the level of the marginal utility of consumption λ. Then the key considerations governing the impact effect of technology on the level of λ are exactly those familiar from the real-business-cycle literature. An improvement in technology raises wealth in the sense of permanent income, tending to lower the marginal utility of consumption λ (through the term $\ln\left(\lambda_\infty\right)$ in (5.2)). Once prices adjust, improved technology also raises the real interest rate as the capital stock adjusts to its new, higher long-run level, tending to raise the marginal utility of consumption, λ, through the integral over future $r - \rho$ in (5.2).

As a benchmark case, suppose the wealth and interest rate effects of the technology change are of equal size and so cancel each other out. Then the Frisch labor supply curve does not shift. A technology improvement then causes both a decline in the quantity of labor L required to produce a given amount of output and a decline in the real wage W. Since W and L are both lower for any given output level, the numerator of equation (5.1) falls, implying a downward shift in the NRR curve.

Even in more complex cases, the NRR curve probably shifts down in response to a technology improvement. Kimball (1998) shows that when technology improves, the NRR curve necessarily shifts down with Cobb-Douglas (or less substitutable) technology and an economic structure that allows steady-state growth—in particular a King-Plosser-Rebelo utility function that makes income and substitution effects on labor supply cancel. Less substitutability than Cobb-Douglas makes the rental rate more sensitive to L/K, strengthening the result. Getting the NRR curve to shift up in response to improved technology requires (a) an effective elasticity of substitution between capital and variable factors greater than 1, (b) a relatively low elasticity of intertemporal substitution to make the interest rate effect small relative to the wealth effect, and (c) a high labor supply elasticity. The elasticity of factor substitution must exceed 1 by enough to overcome both the interest rate effect and the slope of the labor supply curve. Overall, it seems likely that a technology improvement shifts the NRR curve down.

Turning to the monetary policy rule, a reasonable, fairly general specification is that the monetary authority sets the nominal interest rate $r + \pi$ by

$$r + \pi = a + by + g\pi \tag{5.3}$$

where b and h are constants, y is log output, and π is inflation. a evolves over time, with monetary policy tightening if inflation is above its target rate and loosening if inflation is below its target rate.

An LM curve with constant nominal money supply is the case where $g = 0$. For reasonable values of the income and interest rate elasticities of money demand, b would be on the order of 1/year. Ignoring some inessential details, the Taylor (1993) rule fits this description of monetary policy with (roughly) $g = 1.5$ and b also on the order of 1/year. The key difference between an LM curve and the Taylor rule becomes clear when one focuses on the determination of the real interest rate. The LM curve implies

$$r = a + by - \pi ,$$

while the Taylor rule implies

$$r = a + by + 0.5\pi .$$

Thus, when inflation falls temporarily as a result of a technological improvement, an LM curve would shift up, while the Taylor rule curve would shift down. A real interest rate rule is between these cases:

$$r = a + by.$$

Figure 4B shows the upward movement of an LM curve due to the temporary disinflation induced by a technological improvement, interacting with a downward shift in the NRR curve. In this case, output and the real interest rate unambiguously fall in the short run. Note the contrast with the case of a vertical LM curve, discussed in the introduction. In that case output was unchanged, although inputs fell.

Figure 5A shows that with a real interest rate rule, the downward shift of the NRR curve by itself reduces output and the real interest rate in the short-run equilibrium immediately after the shock.

Figure 5B shows that with a Taylor rule, what happens to output (and to the real interest rate) is ambiguous despite the downward shift in the NRR curve. "Sluggish inflation" can resolve the ambiguity. If inflation π does not immediately jump in response to a technological improvement (for reasons that beg for further theoretical exploration), the monetary policy curve immediately after the shock will look like the case of the real interest rate rule in Figure 5A. (Moreover, sluggish inflation delays the adjustment process of the economy towards full employment enough that even q-theory-style investment adjustment costs cannot immediately move the economy off the NRR curve, since those adjustment costs have to interact with motion to have an effect on the real interest rate.) Sluggish inflation appears to be observed empirically (see Fuhrer and Moore, 1995; Roberts, 1998; and Gali and Gertler 1998). Given these considerations, we think Figure 5A probably best captures why a positive technology shock might be contractionary in terms of output as well as in terms of inputs.

Of course, at a firm level, $dy = \mu(dx + du) + dz$, so a fall in output is a stronger result than needed to challenge conventional RBC wisdom. Inputs can fall even when output rises. Nevertheless, our point estimates suggest that output also falls when technology improves, and our discussion in this section shows that such a result is not surprising.

Basu (1998) confirms the bottom line of our theoretical discussion via simulations. He assumes that prices are sticky due to Taylor-style staggered price-setting. He also assumes that monetary policy follows a Taylor rule, and thus responds with a one-period lag to economic conditions. This model, which is similar to the extended Kimball (1995) model that we have discussed, reproduces quite accurately our estimate of the impact effect of a technology improvement. It predicts that output should barely change in the period when technology improves, while inputs should fall significantly.[18] Thus, theoretical investigations show that the results we obtain in Section III are just what one should expect in a world with price stickiness.

B. Evidence On the Sticky-Price Hypothesis

We now explore additional evidence on whether the contractionary effects of technology improvements comes through the sticky-price channel. The sticky-price interpretation works through deflation and real-interest-rate channels, so we first investigate the behavior of those variables. We then study the response of consumption and investment 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 specifying a time-series model for inflation over the post-war period. 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.[19] We measure the price level with the GDP deflator.

The top right panel of Figure 6 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

[18] Since Basu's model does not incorporate many of the "real rigidities" discussed in the literature, it cannot reproduce the prolonged contraction we find in the data. (The lack of a prolonged downturn in his model is also partially due to the assumed Taylor rule, which implies that the monetary authority responds to the fall in inflation by reducing the nominal interest rate.) But Kimball (1995) shows that one can obtain a "contract multiplier" of any desired size by adding real rigidities to the model; it is the impact effect that is significant.

[19] 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.

nominal interest rate returns to its pre-shock value and the Fed does not increase the money supply.)

The middle left panel of Figure 6 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. Figures 4 and 5 show that we expect r to fall initially. We then expect the interest rate to 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.

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 arbitrary time-series process for technology. If consumption and leisure are normal, the fall in labor input with unchanged consumption implies that real wages fall. But real wages and labor input can both fall only if labor demand shifts back. Labor demand depends only on the current level of technology (and the capital stock). In a flexible-price model, even with labor-augmenting technology, improved technology reduces labor demand only if the elasticity of labor demand is less than one. With constant returns to scale, this condition requires the (local) elasticity of substitution between capital and labor to be 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). In any event, we suspect most RBC theorists would not accept parameterizations that imply 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 (1997) document for capital—then technical progress can, in the short run, lead to declines in employment and GDP, as we find.[20] 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 include measures of the dispersion as well as the mean of technology change as explanatory variables in our basic regressions, and see which set of variables has greater explanatory power. A natural dispersion measure, *Disp*, is the cross-sectional standard deviation in technical progress:

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

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.[21]

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 in Table 5, adding the current and two lagged values of *Disp* in addition to current and lagged values of technology impulses, ε. As shown in Table 10, *Disp* has the predicted negative sign on impact, but it is not statistically significant.

Compared with Table 5, the coefficients and standard errors of ε and its lags are not much affected.

---

[20] Lilien (1982) measures reallocative shocks as the cross-industry variance of employment growth rates, and argues for important effects of sectoral shifts. Abraham and Katz (1986) criticize Lilien's measures. Loungani et al. (1990) and Brainard and Cutler (1993) respond to this critique by using the cross-industry dispersion of stock market returns.

[21] 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

The coefficient on $\hat{\varepsilon}_t$ in the fully-corrected case falls to half its previous value, but remains significant. The timing patterns discussed in Section IIIB are unaltered. Most importantly, the addition of the *Disp* variables barely improves the R^2 of the regressions—the increase is between 0.03 and 0.06. 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.[22] For example, Greenwood and Yorukoglu (1996) 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 hence measured 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 also would conclude (incorrectly) that technology fell. Since inputs 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.

---

shifts are important, our index dispersion index should significantly predict input and output growth.

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.[23] 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.

Note that even if cyclical reallocations such as cleansing are important, these effects may not affect the cyclicality of our residuals. 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 $\delta < 0$; 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, equaling (μ +δ), the estimated technology shocks would not suffer from reverse causation. ξ is then a form of classical measurement error, discussed in Section IV. (This could explain why some of our estimated industry markups are less than one.) However, if ξ is correlated with business-cycle variables—

---

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

23 This idea goes back at least to Schumpeter. Foster, Krizan, and Haltiwanger (1998) discuss empirical evidence on the role of entry and exit in aggregate productivity growth.

reallocations may, for example, depend on the aggregate cycle as well as sectoral inputs—then some part of our residuals may remain correlated with output 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 the Granger causality tests discussed in Section IIIB 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 our tests indicate that no such variable 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.

Our model does not make a definite prediction for the behavior of investment on impact (although investment must rise after the period of price adjustment is over). 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 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 be tempted 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.[24] The bottom panels of Figure 6 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. Using various different utilization proxies 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 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

---

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

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. As we show theoretically, if the Fed tries to stabilize the real economy, 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. More work needs to be done to distinguish among the various possible explanations for this fact. Our paper, as well as Gali (1998), advances price stickiness as the major reason for the perverse short-run effect of technical improvement. The evidence is broadly consistent with this view, but it could also be consistent with other models. Two of the main competing explanations are sectoral-shifts models and "cleansing effects" models. (Complicating matters, these hypotheses are not mutually exclusive, but all could contain an element of the truth.) We have presented some evidence that neither is responsible for our findings, but more and sharper tests are needed before we can be sure that price stickiness is in fact responsible for our results. Additional research with firm-level data would be particularly useful in this endeavor.

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?” Forthcoming in Federal Reserve Bank of Boston volume *Beyond Shocks: What Causes Business Cycles?*

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 593, Board of Governors of the Federal Reserve System.

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

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, no4, 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." Paper presented at NBER CRIW Conference, Washington, DC.

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

Gali, Jordi (1998). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, forthcoming.

Gali, 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 (1996). "1974." Rochester Center for Economic Research, Working Paper 429.

Greenwood, Jeremy and Boyan Jovanovic (1998). "Solow vs. Solow: Equipment Prices and Growth." Paper presented at NBER CRIW Conference, Washington, DC.

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.

Henderson, Dale W. and Warwick J. McKibbon (1993). "An Assessment of Some Basic Monetary-Policy Regime Pairs: Analytical and Simulation Results from Simple Multiregion Macroeconomic Models." In Bryant, Hooper, and Mann eds., *Evaluating Policy Regimes: New Research in Empirical Macroeconomics.* Brookings Institution, Washington, DC.

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

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). "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.

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.

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?" Paper presented at the NBER *Macroeconomics Annual* Conference, 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.

Table 1. Descriptive Statistics for Technology Residuals

A. Private Economy

	Mean	**Standard Deviation**	**Minimum**	**Maximum**
Solow Residual	0.011	0.022	-0.044	0.066
Tech. Residual (No Util. Correct.)	0.012	0.016	-0.034	0.050
Tech. Residual (Hours Correct.)	0.013	0.013	-0.013	0.042
Tech. Residual (Full B-K Correct.)	0.013	0.012	-0.013	0.032

B. Manufacturing

	Mean	**Standard Deviation**	**Minimum**	**Maximum**
Solow Residual	0.023	0.035	-0.081	0.080
Tech. Residual (No Util. Correct.)	0.014	0.030	-0.085	0.072
Tech. Residual (Hours Correct.)	0.018	0.028	-0.030	0.082
Tech. Residual (Full B-K Correct.)	0.021	0.021	-0.019	0.067

Note: Sample period is 1950-1989.

Table 2. Basic Data Correlations

A. Private Economy

	Output Growth ***dv***	**Input Growth** dx^V	**Hours Growth** ***dh+dn***	**Solow Residual**
Output Growth ***dv***	1			
Input Growth dx^V	0.78 (0.62, 0.88)	1		
Hours Growth ***dh+dn***	0.80 (0.64, 0.89)	0.91 (0.83, 0.92)	1	
Solow Residual	0.84 (0.72, 0.91)	0.33 (0.02, 0.59)	0.44 (0.15, 0.66)	1

B. Manufacturing

	Output Growth ***dv***	**Input Growth** dx^V	**Hours Growth** ***dh+dn***	**Solow Residual**
Output Growth ***dv***	1			
Input Growth dx^V	0.81 (0.66, 0.90)	1		
Hours Growth ***dh+dn***	0.86 (0.75, 0.92)	0.98 (0.96, 0.99)	1	
Solow Residual	0.84 (0.71, 0.91)	0.36 (0.05, 0.61)	0.46 (0.17, 0.68)	1

Note: 95 Percent confidence intervals in parentheses, calculated using Fisher transformation. Sample period is 1950-1989.

Table 3. Correlations of Technology Residuals with Basic Data

A. Private Economy

	Output Growth ***dv***	**Input Growth** dx^V	**Hours Growth** ***dh+dn***	**Solow Residual**
Tech. Residual (No Util. Correct.)	0.46 (0.17, 0.68)	-0.12 (-0.41, 0.21)	-0.06 (-0.37, 0.26)	0.77 (0.63, 0.88)
Tech. Residual (Hours Correct.)	0.04 (-0.28, 0.35)	-0.42 (-0.65, -0.12)	-0.44 (-0.66, -0.14)	0.40 (0.10, 0.64)
Tech. Residual (Full Correct.)	0.085 (-0.45, 0.26)	-0.37 (-0.62, -0.06)	-0.43 (-0.65, -0.13)	0.42 (0.12, 0.65)

B. Manufacturing

	Output Growth ***dv***	**Input Growth** dx^V	**Hours Growth** ***dh+dn***	**Standard Produc-tivity**
Tech. Residual (No Util. Correct.)	0.42 (0.12, 0.65)	-0.14 (-0.44, 0.18)	-0.04 (-0.35, 0.28)	0.79 (0.63, 0.89)
Tech. Residual (Hours Correct.)	-0.40 (-0.64, 0.10)	-0.64 (-0.80, -0.41)	-0.62 (-0.78, -0.38)	-0.05 (-0.36, 0.27)
Tech. Residual (Full Correct.)	-0.24 (-0.52, 0.08)	-0.51 (-0.71, -0.23)	-0.50 (-0.70, -0.22)	0.09 (-0.23, 0.39)

Note: 95 Percent confidence intervals in parentheses, calculated using Fisher transformation. Sample period is 1950-1989.

Table 4. Variance Decompositions

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. Effect of Technology Improvements on Input Use

	Uncorrected Technology Measure	**Hours-Corrected Technology Measure**	**Fully-Corrected Technology Measure**
$\hat{\varepsilon}_t$	-0.36 (0.16)	-0.69 (0.17)	-0.81 (0.20)
$\hat{\varepsilon}_{t-1}$	0.60 (0.16)	0.24 (0.17)	0.24 (0.20)
$\hat{\varepsilon}_{t-2}$	0.32 (0.16)	0.47 (0.17)	0.44 (0.21)
R^2	0.42	0.45	0.41
D. W.	1.94	1.91	1.87

Note: Dependent variable is aggregate input growth, dx^V. Regressions include a constant. Sample period is 1954-1989.

Table 6. Effect of Technology Improvements on Utilization

	Hours-Corrected Technology Measure	**Fully-Corrected Technology Measure**
$\hat{\varepsilon}_t$	-0.23 (0.12)	-0.44 (0.18)
$\hat{\varepsilon}_{t-1}$	0.35 (0.12)	0.48 (0.18)
$\hat{\varepsilon}_{t-2}$	0.08 (0.12)	0.27 (0.18)
R^2	0.28	0.33
D. W.	2.06	2.23

Note: Dependent variable is aggregate utilization change. Regressions include a constant. Sample period is 1954-1989.

Table 7. Effect of Technology Improvements on Output Growth

	Uncorrected Technology Measure	**Hours-Corrected Technology Measure**	**Fully-Corrected Technology Measure**
$\hat{\varepsilon}_t$	0.55 (0.26)	-0.29 (0.33)	-0.39 (0.38)
$\hat{\varepsilon}_{t-1}$	1.29 (0.27)	1.13 (0.33)	1.28 (0.38)
$\hat{\varepsilon}_{t-2}$	0.17 (0.27)	0.79 (0.34)	0.82 (0.38)
R^2	0.47	0.35	0.35
D. W.	2.05	2.07	2.12

Note: Dependent variable is aggregate output growth, dv. Regressions include a constant. Sample period is 1954-1989.

Table 8. Correlations Using Electricity-Corrected Residuals

A. Electricity-Corrected Regression Residuals

	Output Growth dv	**Input Growth** dx^V	**Hours Growth** $dh+dn$	**Standard Productivity**
Tech. Residual (Private Economy)	0.41 (0.11, 0.64)	-0.13 (-0.43, 0.19)	-0.08 (-0.39, 0.24)	0.71 (0.51, 0.84)
Tech. Residual (Manufact.)	0.23 (-0.09, 0.51)	-0.25 (-0.52, 0.07)	-0.20 (-0.49, 0.12)	0.60 (0.35, 0.77)

B. Electricity-Corrected Solow Residuals

	Output Growth dv	**Input Growth** dx^V	**Hours Growth** $dh+dn$	**Standard Productivity**
Solow Residual (Private Economy)	0.45 (0.16, 0.67)	0.02 (-0.30, 0.33)	0.04 (-0.28, 0.35)	0.64 (0.41, 0.80)
Solow Residual (Manufact.)	0.50 (0.22, 0.70)	-0.04 (-0.35, 0.28)	0.03 (-0.29, 0.34)	0.82 (0.68, 0.90)

Note: 95 Percent confidence intervals in parentheses, calculated using Fisher transformation. Sample period is 1950-1989.

Table 9. One-Digit and Industry-Average Correlations

	Corr(*dp*,*dy*)	Corr(*dp*, *dx*)	Corr(*dz*, *dy*)	Corr(*dz*, *dx*)
Mining	0.33*	-0.77*	0.91*	0.06
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 33 Industries (21 Manufact, 11 other)	0.53	0.03	0.33	-0.10

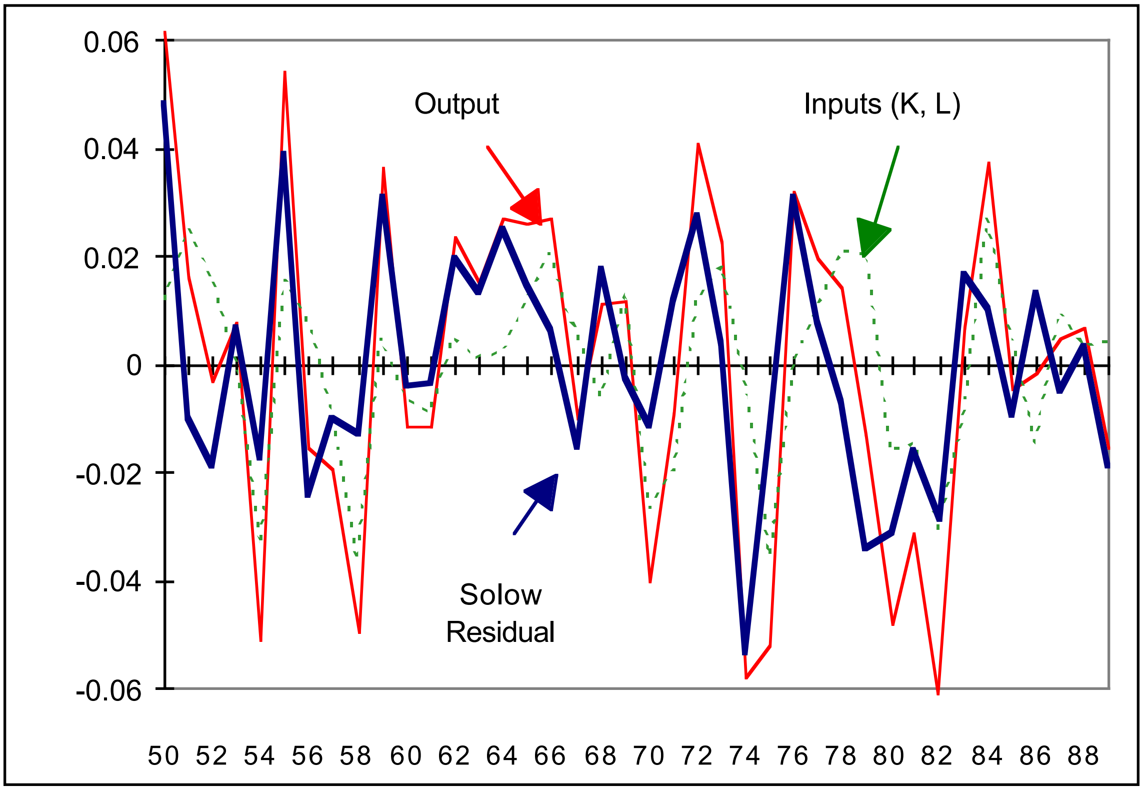
Note: The 33 individual industries and the 10 one-digit industries span the private non-farm business economy. Except where noted, all averages are output weighted. Technology *dz* is calculated as the residual from an industry-by-industry gross-output regression as described in Section II, with the 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 10. Effect of Technology Dispersion on Input Growth

	Uncorrected Technology Measure	**Hours-Corrected Technology Measure**	**Fully-Corrected Technology Measure**
$\hat{\varepsilon}_t$	-0.40 (0.16)	-0.71 (0.17)	-0.44 (0.20)
$\hat{\varepsilon}_{t-1}$	0.61 (0.16)	0.19 (0.17)	0.17 (0.21)
$\hat{\varepsilon}_{t-2}$	0.31 (0.16)	0.52 (0.17)	0.44 (0.21)
$Disp_t$	-0.09 (0.22)	-0.15 (0.18)	-0.32 (0.25)
$Disp_{t-1}$	0.25 (0.23)	0.24 (0.19)	0.24 (0.26)
$Disp_{t-2}$	-0.28 (0.22)	-0.37 (0.19)	-0.30 (0.25)
R^2	0.45	0.51	0.46
D. W.	1.92	2.00	2.04

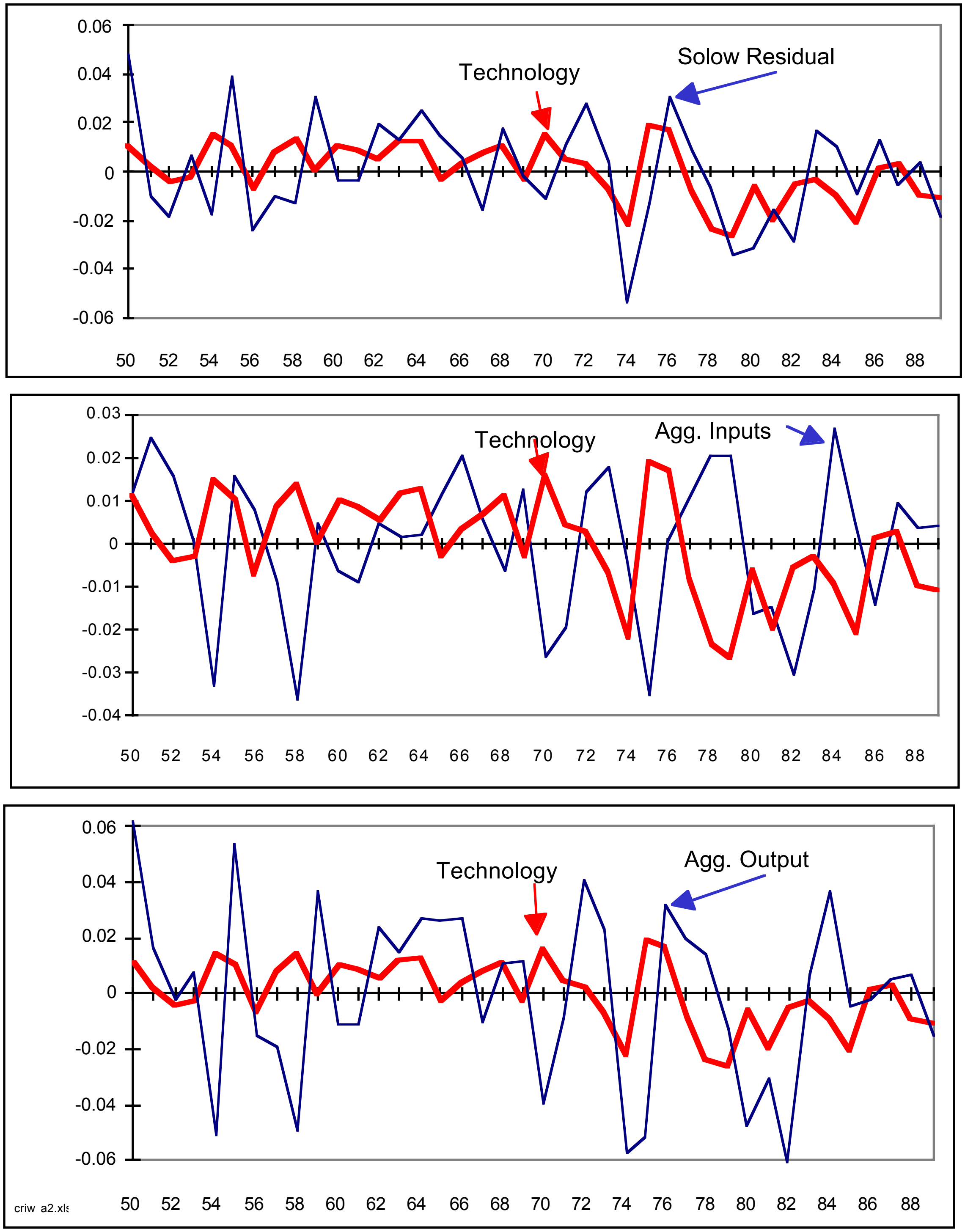
Note: Dependent variable is aggregate input growth, dx^V. Regressions include a constant. Sample period is 1954-1989.

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



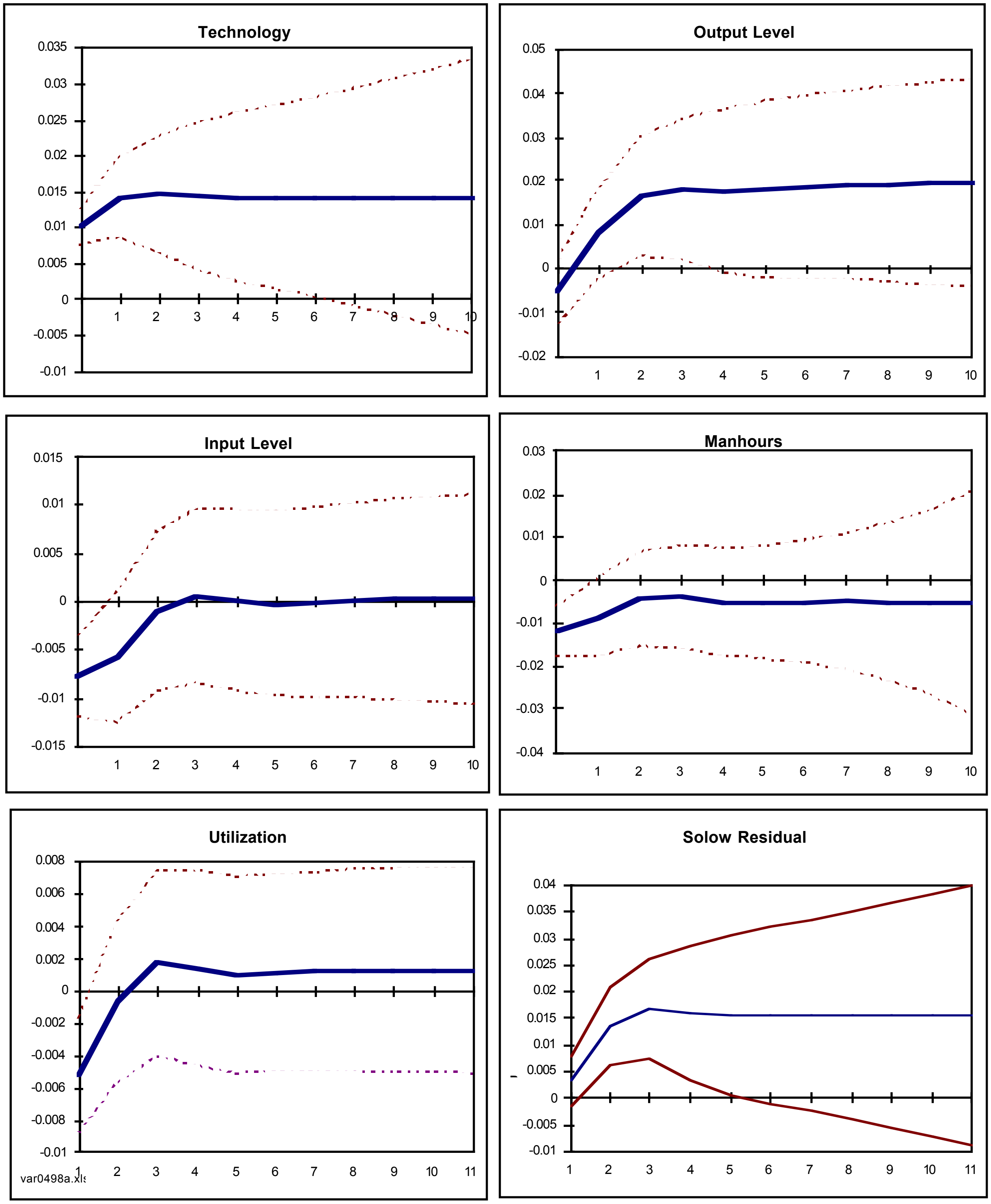
Note: All series are demeaned. Sample period in 1950-89.

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



Note: The technology series is the fully-adjusted residual. All series are demeaned. Entries are percent changes. Sample period is 1950-89.

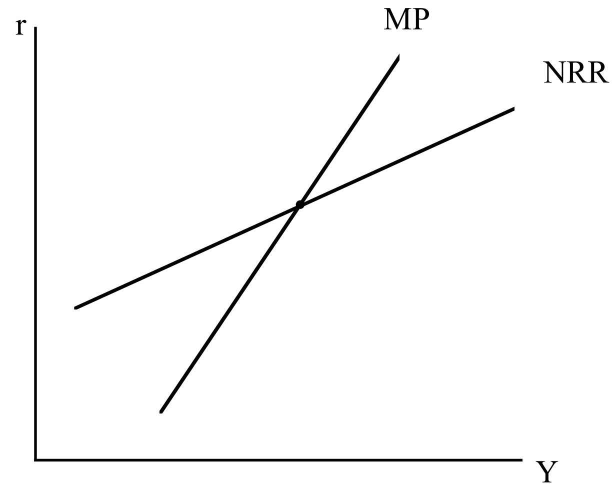
Figure 3. Impulse Responses to Technology Improvement: Basic Variables



Note: Impulse responses to a 1 percent (0.01) improvement in technology. The technology series is the fully-adjusted 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

A. Short-Run Equilibrium with Sticky Prices



B. Technology Improvement with Money Held Constant

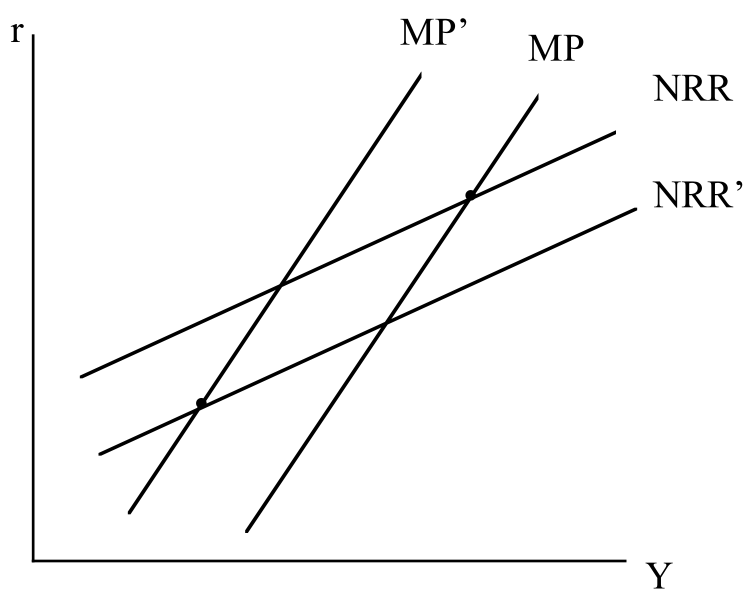
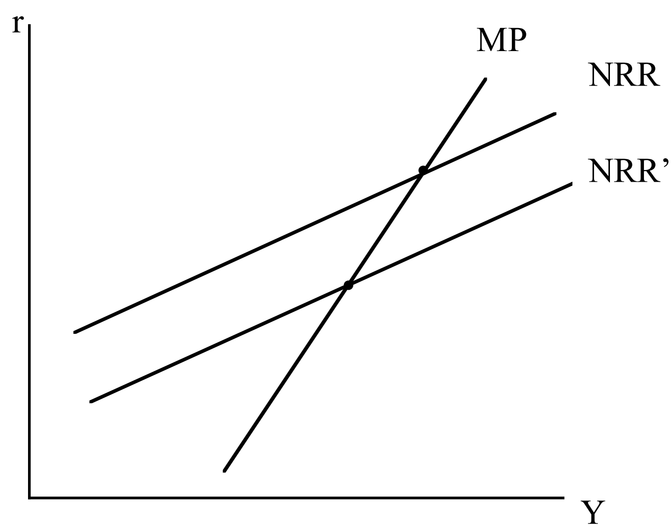


Figure 5

A. Technology Improvement with a Real Interest Rate Rule



B. Technology Improvement with a Taylor Rule and Jumping Inflation

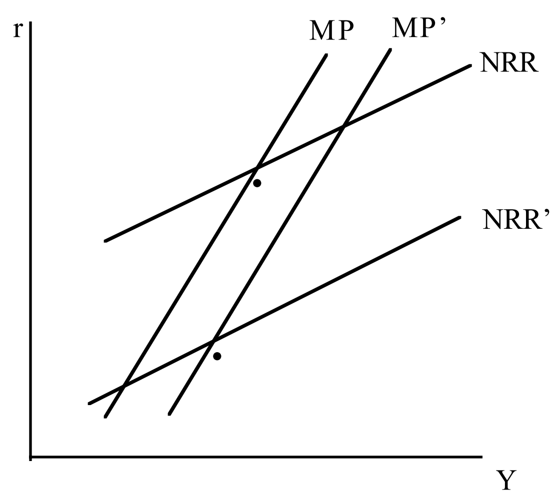
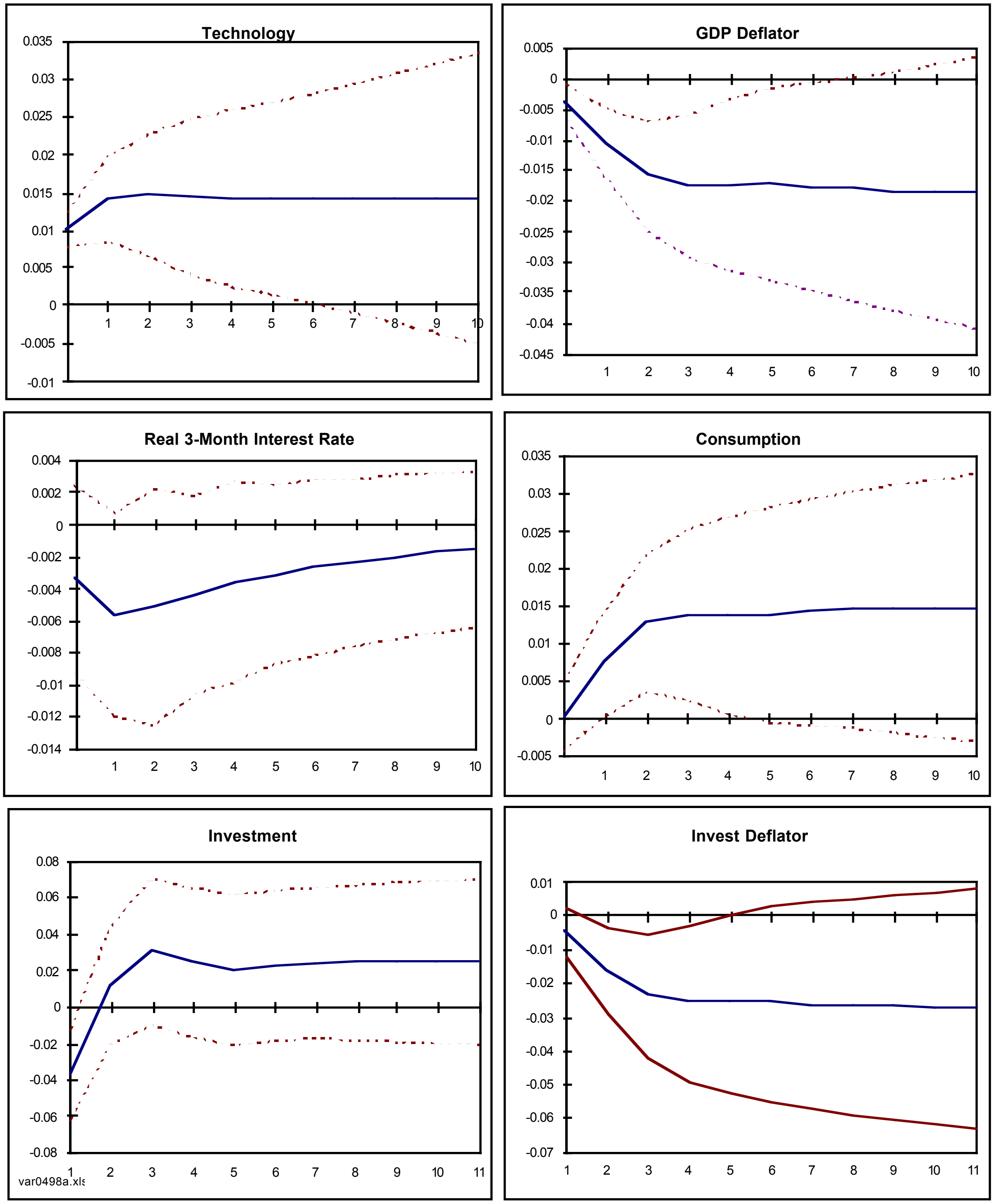


Figure 6. Impulse Responses to Technology Improvement: Other Variables



Note: Impulse responses to a 1 percent (0.01) improvement in technology. The technology series is the fully-adjusted residual. All units are percent, except for the real interest rate, which is in percentage points. Dotted lines show 95 percent confidence intervals, computed using Monte Carlo bootstrap method. Sample period is 1952-89.