

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 593

October 1997

AGGREGATE PRODUCTIVITY AND AGGREGATE TECHNOLOGY

Susanto Basu and John G. Fernald

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.bog.frb.fed.us.

AGGREGATE PRODUCTIVITY AND AGGREGATE TECHNOLOGY

Susanto Basu and John G. Fernald*

Abstract: Aggregate productivity and aggregate technology are meaningful but distinct concepts. We show that a slightly-modified Solow productivity residual measures changes in economic welfare, even when productivity and technology differ because of distortions such as imperfect competition. We then present a general accounting framework that identifies several new non-technological gaps between productivity and technology, gaps reflecting imperfections and frictions in output and factor markets. Empirically, we find that these gaps are important, even though we abstract from variations in factor utilization and estimate only small average sectoral markups. Compared with productivity growth, our measured technology shocks are significantly less correlated with output, and are essentially uncorrelated with inputs. Our results imply that calibrating dynamic general equilibrium models as if Solow residuals were technology shocks confuses impulses and propagation mechanisms.

Keywords: productivity, technology, aggregation, welfare

* Fernald is an economist in the International Finance Division of the Federal Reserve Board, and can be contacted at fernaldj@frb.gov. Basu is an assistant professor of economics at the University of Michigan, a visiting assistant professor of economics at Harvard University, and a faculty research fellow at the National Bureau of Economic Research; he can be contacted at sbasu@umich.edu. A substantially different version of this paper was circulated under the title “Aggregate Productivity and the Productivity of Aggregates.” We thank Martin Eichenbaum, Robert Hall, John Haltiwanger, Michael Horvath, and Miles Kimball for valuable suggestions, and participants in a number of seminars for insightful comments. Basu gratefully acknowledges a National Fellowship at the Hoover Institution, where the first draft of this paper was written, financial support from the National Science Foundation, and a Research Fellowship from the Alfred P. Sloan Foundation. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

In an economy without distortions, Solow's (1957) productivity residual indexes aggregate technology change.¹ But what if productivity and technology differ because of distortions such as imperfect competition? Recent macroeconomic literature often seems to assume that we measure productivity in order to measure technology; any differences reflect "mismeasurement."²

Some authors suggest ways to fix this perceived measurement problem. For example, Hall (1988, 1990) shows how to estimate technology change for imperfectly competitive firms, a method that extends to an aggregate level if there is a representative firm.³ If no representative firm exists, however, it remains unclear how to measure aggregate productivity and aggregate technology.

In this paper, we undertake three tasks. First, we clarify the meaning and measurement of aggregate productivity. Productivity has clear welfare implications, even in a world with distortions. A modified Solow residual — which reduces to Solow's measure if there are no economic profits — approximates to first order the welfare change of a representative consumer. Intuitively, growth in aggregate output measures the growth in society's ability to consume. To measure welfare change, we must then subtract the opportunity cost of the inputs used to produce this output growth. Input prices measure that cost, regardless of whether they also reflect marginal products. Hence, if productivity and technology differ, then it is productivity that most closely indexes welfare.⁴

Second, we relate aggregate productivity to aggregate technology. We provide a general accounting framework that shows how imperfections and frictions in output and factor markets matter.

¹ Solow assumed there was an aggregate production function; Hulten (1978) shows that Solow's result also holds if one computes technology change from disaggregated data and then aggregates appropriately.

² The growth-accounting literature does not always take this view. For example, Denison (1979) views productivity as measuring the ability of society to increase standards of living.

³ Note, however, that even if all firms have identical technology and behavior, a representative firm may not exist (even to a log-linear approximation). Additional strong assumptions are needed: e.g. perfect factor mobility, and a zero elasticity of substitution in production between primary and intermediate inputs. We discuss this issue in Section III, below; see also Basu (1995) and Rotemberg and Woodford (1995).

⁴ Throughout the paper, we assume that inputs and outputs are correctly measured. We thus abstract from cyclical mismeasurement (such as variations in work effort or capital utilization), which in practice probably contribute to

For example, if firms have different markups of price over marginal cost, or pay different wages, then society may value resources differently in different uses. Reallocating resources towards highly valued uses raises aggregate productivity, without necessarily reflecting changes in technology. Because aggregate data include these non-technological reallocations, measuring technology requires disaggregated data; by contrast, the modified Solow productivity residual requires only aggregate data.

Third, we explore empirically the differences between aggregate productivity growth and aggregate technology growth. We use data from 34 roughly two-digit industries that together constitute the U.S. private business economy, and estimate the non-technological terms from our accounting identity.⁵ We find that compared with productivity growth, our measured technology shocks are significantly less correlated with output. More importantly, the technology shocks are essentially uncorrelated with inputs.

Recent empirical evidence suggests that the average degree of sectoral market power is small.⁶ It may thus seem surprising that we find nevertheless that imperfect competition has an important influence on aggregate productivity. For example, Hall (1988, 1990) emphasizes that with imperfect competition, productivity rises when primary inputs of capital and labor increase. This channel depends on the existence of a substantial gap between price and marginal cost at a representative firm.

However, imperfect competition affects productivity through two channels in addition to Hall's "average" effect. First, a factor's marginal product may differ across firms: firms may have different degrees of market power in output markets, or pay different prices for the same input. For example, suppose durable-goods firms have larger markups than non-durable-goods firms. Since durable-industries are more cyclical, they employ a larger share of the marginal inputs in a boom. This marginal

differences between productivity and welfare. We do so to keep the issues conceptually clean: the distortions on which we focus reflect true, not spurious, differences between correct measures of productivity and technology.

⁵ Ideally, our decomposition requires firm-level data. However, no firm-level datasets exist that cover the entire economy. Since we focus on aggregate productivity and technology, we use industry data; we discuss this choice in Section IV. Our work complements firm-level studies of narrowly-defined industries (e.g., Bertin et al. (1995)).

⁶ For example, Burnside (1996) and Basu and Fernald (1997). These authors reports estimates of returns to scale, but given that rates of pure profit appear small, their results also provide evidence on the size of markups.

reallocation makes productivity cyclical even if technology does not change, and even if the average sector has only small markups. Second, with imperfect competition, intermediate-input use in general affects aggregate value added. Value added takes gross output and subtracts intermediate inputs valued at their purchase price, not their marginal product. Imperfect competition drives a wedge between the price and the marginal product of these intermediates, a wedge that represents real goods and services. These two channels — which depend on the allocation of resources across uses — turn out to contribute to aggregate productivity. Thus, we identify several new gaps between productivity and technology.

Our algebra incorporates and clarifies the role of existing stories for why productivity and technology differ. For example, in the sectoral shifts literature (e.g. Phelan and Trejos (1996)), demand shocks cause differences in the marginal product of immobile factors across firms. Also, empirical microeconomists note that within narrowly defined industries, productivity change often reflects reallocations of resources among firms with different levels of productivity (e.g. Baily et al. (1992)).

A central conclusion of our paper is that aggregate productivity and aggregate technology are meaningful but distinct concepts. When productivity and technology differ, the differences are not “biases”: they directly affect welfare, since they reflect real goods to consume or invest.

This conclusion matters for distinguishing the impulses driving business cycles from the mechanisms propagating those impulses. Real business cycle (RBC) models use Solow’s productivity residual — interpreted as technology shocks — as the dominant impulse driving the cycle (e.g. Cooley and Prescott (1995)). Other impulses may affect output in these models, but technology shocks must dominate to match the key stylized fact of business cycles: the positive comovement of output and labor input.⁷

⁷ Barro and King (1984) provide an early discussion of this issue. Dynamic general equilibrium models without technology shocks can match this stylized fact with countercyclical markups of price over marginal cost, arising from sticky prices (as in Kimball (1995)) or from game-theoretic firm interactions (as in Rotemberg and Woodford (1992)). Models with an extreme form of increasing returns — increasing marginal product of labor — can also produce a positive comovement between output and labor input; see, e.g., Farmer and Guo (1994).

If firms are not all perfectly competitive, then using the Solow residual as a measure of technology shocks mixes impulses and propagation mechanisms. Technology change is the impulse; sectoral reallocations induced by technology or other shocks are propagation mechanisms. We need to measure the two separately. In principle, reallocations can either increase or reduce productivity. Even if the apparent size of technology shocks falls, these shocks may nevertheless remain important for explaining business cycles, since reallocations are a new propagation mechanism. In theory, the reallocations induced by technology shocks could increase the fraction of output volatility that they explain.

However, modeling this propagation mechanism formally requires a multi-sector dynamic general equilibrium model. Thus, our empirical results serve two purposes. First, they provide a better technology change series than those currently extant, and thus better measure the impulses coming from technology change. Second, they suggest that one-sector business-cycle models miss important propagation mechanisms, and thus may guide the construction of new models.

The paper comprises six sections. Section I presents basic definitions. Section II proves the welfare properties of our new definition of productivity. Section III shows the difference between aggregate productivity growth and aggregate technology change, and suggests how we might estimate the latter. Section IV discusses our data and empirical method. Section V presents our empirical results, and Section VI summarizes our conclusions and indicates directions for future research.

I. Definitions

In the next section, we examine the welfare of a representative consumer, who provides inputs of capital and labor to firms in the economy and then uses his or her income to purchase final goods to consume or invest. We define aggregate variables to be consistent with this focus.

There are N goods in the economy. Firms, indexed by i , produce goods by hiring labor, L_i , and renting capital, K_i , from the representative consumer. We defer discussion of firm-level technology and

behavior until Section III, since they do not affect our results on welfare and productivity. We assume, however, that only one firm produces each good.⁸ When there is no ambiguity, we omit time subscripts.

Since the representative household provides the labor and capital, we define aggregate inputs as simple sums of the firm-level quantities:

$$K \equiv \sum_{i=1}^N K_i,$$

$$L \equiv \sum_{i=1}^N L_i.$$

For simplicity, we assume that there is one type of capital and one type of labor. With heterogeneous capital and labor, we would sum over each input separately; nothing fundamental would change.

In principle, different firms may pay different prices for a homogeneous input. For any input J , let P_{Ji} be the price it pays to rent or hire the input for one period. We define the aggregate (rental) prices of capital and labor as the factor payments divided by aggregate quantities:

$$P_K \equiv \frac{\sum_{i=1}^N P_{Ki} K_i}{K},$$

$$P_L \equiv \frac{\sum_{i=1}^N P_{Li} L_i}{L}.$$

We use the standard Divisia definition of aggregate output, which weights goods by market prices and hence avoids substitution bias in the aggregate output and price indices. Divisia aggregates are defined most naturally in growth rates, and we denote the growth in aggregate output (equivalently, aggregate value added) by dv . The national accounting identity tells us that we can define aggregate output by looking at either production (aggregating value added over firms) or final expenditure (aggregating sales for consumption, investment, government purchases, or export). We will use both approaches; welfare (in Section II) relates to the expenditure side, while production (in Section III) relates to the value-added side.

⁸ This is not a restrictive assumption, since we can index goods by the firms that produce them. Consumers may regard two goods as perfect substitutes, implying that they have identical prices. More restrictive is our implicit

We assume for simplicity that the economy is closed. The consumer buys the final goods to consume or invest; we take government purchases as a form of consumption by the representative household. We also assume that only the N th good is used as capital (though our results easily generalize to having many capital goods). P_i is the price of good i , and C_i is its quantity, where $i = 1, \dots, N - 1$. P^I is the price of the N th (investment) good and I is its quantity. The growth rate of the Divisia index of aggregate output is a weighted average of the growth rates of each component of expenditure. We write this as:

$$(1) \quad dv \equiv \frac{P^C C}{P^V V} \left[\sum_{i=1}^{N-1} \frac{P_i C_i}{P^C C} \frac{dC_i}{C_i} \right] + \frac{P^I I}{P^V V} \frac{dI}{I}$$

where $P^C C = \sum_{i=1}^{N-1} P_i C_i$ is nominal expenditure on consumption goods, $P^I I$ is nominal expenditure on the investment good (good N), and $P^V V = P^C C + P^I I$ is total nominal output. This definition of dv is the continuous-time version of the chain-linked measure of national output produced in the National Income and Product Accounts. The national accounts identity ensures that national output calculated from the expenditure side (as we have done here) equals the Divisia aggregate of firm-level value added.

Investment determines the accumulation of capital:

$$(2) \quad K_t = (1 - \delta)K_{t-1} + I_t.$$

Finally, we define aggregate inputs and aggregate productivity in the ways we use them throughout the paper. We define the share of each input J in aggregate value added as $s_J^V = P_J J / P^V V$. Aggregate inputs, dx^V , are then a share-weighted sum of primary inputs of capital and labor:

$$(3) \quad dx^V = s_K^V dk + s_L^V dl.$$

The shares are the aggregate *cost* of each input divided by total nominal *revenue*; note that in contrast to Solow, we do not take capital's share as a residual. Thus, s_K^V and s_L^V sum to less than one if firms make

assumption that there is a fixed number of firms.

economic profits. We define aggregate productivity growth as output growth minus aggregate input growth:

$$(4) \quad dp = dv - dx^V.$$

If there are economic profits, then our measure of productivity differs from Solow's because input weights do not sum to one. If economic profits are zero, then our weights and hence our productivity measure match Solow's. Given that pure profits appear small (Basu and Fernald (1997); Rotemberg and Woodford (1995)), in practice our measure is very close to Solow's.

II. Productivity and Welfare

Why is aggregate productivity growth interesting? The usual justification is Solow's (1957) proof that with constant returns to scale, perfect competition, and no frictions, it measures aggregate technology change. But in a world with distortions, is the Solow residual merely mismeasured technology?

In this section we show that productivity growth computed from aggregate data has a natural welfare interpretation, regardless of whether it also measures technology change. In particular, the modified Solow residual defined in Section I — which reduces to Solow's residual if there are no economic profits — turns out to measure welfare change for a representative consumer. This result holds even with imperfect competition in output markets and non-constant returns to scale in production.

This conclusion is appealing for two reasons. First, it shows that productivity rather than, say, GDP, is the right measure of economic welfare under fairly general conditions. Second, it shows that even with distortions, policymakers can compute interesting quantities from aggregate data —we do not always need to calculate firm-level or aggregate technology change. As we discuss in the next section, in the short to medium run productivity can change for reasons unrelated to technology change.

Suppose a large number of identical households solve the following intertemporal problem:

$$\text{Maximize } U = \sum_{s=0}^{\infty} \beta^s u(C_{1,t+s}, C_{2,t+s}, \dots, C_{N-1,t+s}, \bar{L} - L_{t+s})$$

(5) subject to

$$A_{t+1} = A_t + P_{L_t} L_t + P_{K_t} K_t - \delta P_{t+1}^I K_t + (P_{t+1}^I - P_t^I) K_t + r_t B_t + \Pi_t - \sum_{i=1}^{N-1} P_{i,t} C_{i,t}.$$

As before, K is capital, L is labor, P^I is the price of investment goods, P_{K_t} is the rental price of capital, and P_{L_t} is the real wage. For now, we assume that all producers pay homogenous inputs the same factor price: that is, for any input J and firm i , $P_{Jit} = P_{Jt}$ (we discuss this assumption further below and in Section III.E). \bar{L} is each consumer's per-period endowment of labor. In the budget constraint, household assets, A , are the sum of private consumption bonds, B (whose price we normalize to 1), and capital, which depreciates at rate δ . The rate of return on bonds is r . Any pure profits, Π , are rebated lump-sum to consumers. In equilibrium, $B_t = 0$ for all t and $(1 + r_t) P_{K_t} = (1 + r_t) P_t^I - (1 - \delta) P_{t+1}^I$.⁹

Let λ_t be the shadow value of assets at time t . The consumer's first-order conditions are:

$$(6) \quad \begin{aligned} u_{L_t} &= \lambda_t w_t, \\ u_{C_{i,t}} &= \lambda_t P_{i,t} \quad i = 1, K, N-1, \end{aligned}$$

along with the Euler equation:

$$(7) \quad \lambda_t = \beta(1 + r_t) \lambda_{t+1}.$$

Suppose a temporary (one-period) shock at time t potentially changes output, consumption, and labor supply. This shock may represent a change in technology, market structure (markups of price over marginal cost), or government expenditure (which we have not formally modeled). We require, however, that the change last only one period; the shock affects outcomes beyond period t only by changing the capital stock.¹⁰ To a first approximation, this perturbation changes lifetime utility by:

⁹ Our timing conventions differ from those of Hall and Jorgenson (1967), who assume that capital is paid in the period after it is rented. By contrast, we assume that capital is paid in the same period that it is rented, matching the convention used in the empirical literature on cyclical productivity (e.g. Hall (1990)).

¹⁰ Our result extends easily to multi-period perturbations: essentially, one then adds up the discounted values of productivity change over as many periods as the shock lasts (which can be infinite). We are implementing this

$$\begin{aligned}
(8) \quad dU &= \sum_{i=1}^{N-1} u_{C_{i,t}} dC_{i,t} - u_{L_t} dL_t + \beta \lambda_{t+1} dA_{t+1} \\
&= \lambda_t \left[\sum_{i=1}^{N-1} P_{i,t} dC_{i,t} - P_{L,t} dL_t \right] + \beta \lambda_{t+1} dA_{t+1}.
\end{aligned}$$

From the national income identity (equation (1)) we can express the change in the value of consumption in terms of changes in aggregate output and investment:

$$\begin{aligned}
(9) \quad \sum_{i=1}^{N-1} P_{i,t} dC_{i,t} &= P_t^C C_t \left[\sum_{i=1}^{N-1} \frac{P_{i,t} C_{i,t}}{P_t^C C_t} \frac{dC_{i,t}}{C_{i,t}} \right] \\
&= P_t^V V_t \frac{dV_t}{V_t} - P_t^I dI_t.
\end{aligned}$$

Since the capital stock at time $t-1$ is predetermined when the shock occurs at time t , the change in investment at time t is also the change in the time- t capital stock: $dI_t = dK_t$.

Using the previous result and the Euler equation (7), and noting that the only asset in this economy is capital, we can rewrite equation (8) as:

$$\begin{aligned}
(10) \quad dU &= \lambda_t \left[P_t^V V_t \frac{dV_t}{V_t} - P_t^I dK_t - P_{L,t} dL_t \right] + \beta \lambda_{t+1} dA_{t+1} \\
&= \lambda_t \left[P_t^V V_t \frac{dV_t}{V_t} - P_t^I dK_t - P_{L,t} dL_t + \frac{(1-\delta)P_{t+1}^I}{1+r_t} dK_t \right].
\end{aligned}$$

We can simplify this expression by pulling nominal output, $P^V V$, outside the brackets and rearranging to put the equation in logarithmic form:

$$\begin{aligned}
(11) \quad dU &= (\lambda_t P_t^V V_t) \left[dv_t - \frac{P_{K,t} K_t}{P_t^V V_t} dk_t - \frac{P_{L,t} L_t}{P_t^V V_t} dl_t \right] \\
&= (\lambda_t P_t^V V_t) dp_t.
\end{aligned}$$

Equation (11) says that the change in utility is proportional to the change in the modified Solow residual dp_t . Thus, the welfare change of any one-period shock at time t is equivalent to having an exogenous

increase in national income of dp percent at time t , which is then added to the consumer's assets.¹¹ Note, however, that the welfare benefit is proportional to productivity growth, not to output growth, since the consumer subtracts the welfare cost of supplying any extra capital and labor.

Since economic profits appear small, the modified residual dp is very close to the standard Solow residual. Thus, even with distortions such as imperfect competition, when the aggregate Solow residual does not in general index technology change, it remains an excellent index of welfare change. Hence, it remains an appropriate target for policy, as well as a convenient indicator.

In the next section, we show that aggregate technology change needs to be measured using disaggregated — ideally, firm-level — data. So why do aggregate data yield a meaningful measure of welfare change? The welfare properties of the Solow residual follow from the equality of relative market prices to the consumer's marginal rates of substitution (MRS) between goods; this includes the equality of the real wage to the MRS between goods and leisure. These equalities hold *even when market prices do not reflect the economy's marginal rate of transformation (MRT) between those goods*. Equivalently, we need only to investigate the expenditure side of the National Income Accounts identity; we do not need to know the production technology of firms or the competitive structure of industries.

There are two qualifications to our argument. First, the ratio of factor prices may not equal the consumer's marginal rate of substitution: taxes, for example, tend to create a wedge between the two, since the wage perceived by firms then differs from the wage received by households. The welfare interpretation of the residual requires factor prices as perceived by the household, but this modification is straightforward: all prices should be those perceived by the household. Second and more seriously, the representative-consumer assumption may fail. Consumers may have different marginal utilities of wealth or, as in standard efficiency-wage models or bargaining models, they may face different prices. In this case, one cannot compute aggregate welfare change from aggregate statistics alone. However, we

¹¹ Since we can normalize the price of value added to 1 without loss of generality, one should think of the welfare

do not claim that our proposed productivity measure is a completely general measure of welfare change, merely that it is one under much more general conditions than the usual Solow residual. It seems a particularly apt measure in the context of recent macroeconomic models with a representative consumer but with imperfect competition in product markets (e.g., Rotemberg and Woodford (1992, 1995)), or with multiple sectors and costly factor reallocation (e.g., Ramey and Shapiro (1997)).

Figure 1 shows the economic intuition underlying our argument. Suppose the economy produces two goods, both of which are consumed (and possibly also used as intermediate inputs). To keep the graph simple, assume that the supplies of capital and labor are fixed. The production possibilities frontier (PPF) depicts all feasible (C_1, C_2) pairs. An economy without distortions attains the social optimum at point A , supported by relative prices P_1/P_2 . Now suppose there are distortions. Then the economy might be at an allocatively inefficient point on the PPF, like point B , or even within the PPF, like point C . As shown in Figure 1, these outcomes can be supported by price ratios different from the MRT between C_1 and C_2 .

Note that in all cases the consumer's budget line shows the economy's iso-output line, which aggregates heterogeneous output using market prices (regardless of whether these prices reflect technological tradeoffs).¹² Thus, in this example welfare increases only if output increases. (This is a special case of our general result that welfare increases only if productivity increases, since in this example $dk = dl = 0$.) Hulten (1978) shows that under Solow's conditions — perfect competition and constant returns — aggregate productivity growth represents both technology improvement and welfare increase. In terms of Figure 1, Hulten's result applies to an economy at point A : Output (productivity) can increase only if the PPF shifts out at point A , i.e. if there is (local) technological improvement.

change as proportional to a *real* change in national income.

¹² For a proof, differentiate the consumer's budget constraint holding income and prices fixed.

However, the same is not true at points *B* and *C*: Output (productivity) can increase without any change in technology, as long as distortions lessen. But these productivity improvements raise welfare, since output and inputs are weighted using prices that reflect the consumer's MRS between goods. Thus, our finding generalizes Hulten's (1978) result to the case of imperfect competition and non-constant returns, and clarifies the essence of his argument linking productivity and welfare.¹³

III. Productivity and Technology

What determines aggregate productivity? In addition to aggregate technology, we identify three non-technological influences, reflecting imperfections and frictions in output and factor markets. First, as Hall (1988, 1990) notes, with imperfect competition, productivity rises when primary inputs of capital and labor rise. This channel requires a substantial gap between price and marginal cost at the representative firm. Second, factors may have different marginal products in different uses: firms may have different degrees of market power in output markets, or face different prices for the same input. Third, changes in intermediate-input use can affect aggregate output growth. Value added subtracts intermediate inputs valued at their purchase price rather than their marginal product. With imperfect competition, the marginal product exceeds the purchase price, creating a wedge representing real goods and services.

We begin with firm-level production functions for gross output, and then derive the value-added analogues. At a disaggregated level, the natural measure of firm output is gross output. Real value added is an artificial construct — bread without flour; books without paper or ink; shoes lacking leather.

We focus on firm-level value added, however, because of our ultimate interest in aggregates.

Aggregate final expenditure — private and public consumption, investment, and net exports — measures

¹³ While our result is related to Hulten's, our method is more closely related to Weitzman's (1976) argument that Net National Product is a better national income accounting concept than Gross National Product. Indeed, Weitzman's model is a special case of ours: It assumes that period utility is linear in consumption, that future utility is discounted

what society consumes today or saves for tomorrow. The national accounts identity shows that aggregate final expenditure equals the aggregate of firm-level value added — intermediate-input use cancels out. Thus, aggregating firm-level value added leads to economically sensible aggregates.

We abstract from issues of mismeasurement, such as variable capital utilization or labor hoarding. Mismeasurement is probably important in the data, but is conceptually distinct from the issues we discuss. Properly measured, aggregate productivity growth increases welfare regardless of its source. However, if productivity growth reflects input-mismeasurement alone, it does not increase welfare. That is, in a world without frictions, consumers are indifferent at the margin between supplying extra inputs — observed or unobserved — and consuming the extra output that these inputs produce.

Subsection A analyzes the determinants of firm-level value added and productivity. Subsection B uses those microfoundations to analyze aggregate output and productivity. The final three subsections explore different economic perspectives on our results.

A. The Firm Level

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

$$(12) \quad Y_i = F^i(K_i, L_i, M_i, T_i),$$

where Y_i is gross output, K_i , L_i , and M_i are inputs of capital, labor, and materials, and T_i indexes technology. The firm's production function F^i may be homogeneous of arbitrary degree γ_i in K_i , L_i , and M_i . γ_i is not constrained to be one, so F^i may have non-constant returns to scale.

Note that increasing returns may take different forms. For example, firms may have fixed costs of production along with constant or increasing marginal cost, as in Rotemberg and Woodford (1995); alternatively, firms may have no fixed costs but declining marginal cost, as in Farmer and Guo (1994). Rotemberg and Woodford (1995) emphasize that the source of increasing returns matters for how shocks

at the interest rate, and that consumers do not value leisure.

affect macroeconomic models, and argue that fixed costs are the most plausible explanation of increasing returns at a firm level. However, the source of increasing returns does not affect the derivations below, where we present an accounting identity for output change at each moment in time.

We assume that firms are price takers in factor markets, but may have market power in output markets. For any input J , let F_J^i be the marginal product. Firm i 's first-order conditions then imply that the value of the marginal product is proportional to the price it pays for the input, P_{ji} :¹⁴

$$(13) \quad P_i F_J^i = \mu_i P_{ji}.$$

Firms may charge a markup, μ_i , over marginal cost: $\mu_i = P_i / MC_i$, where MC_i is marginal cost.

By definition, returns to scale γ_i equals the sum of the output elasticities with respect to all inputs. Combining this with the first-order conditions, it is straightforward to find the relationship among μ_i , γ_i , and the ratio of economic profit to total revenue, s_π^i :

$$(14) \quad \gamma_i = \mu_i (1 - s_\pi^i).$$

Below, we emphasize markups rather than returns to scale, since the markup determines how the social and private valuations (i.e., the marginal product and the input price) of a factor differ. However, the equation above shows the close relationship between returns to scale and imperfect competition: firms with increasing returns must charge markups to cover their costs. If economic profits are small, μ_i and γ_i must be approximately the same. If returns to scale differ across firms for technological reasons (such as differences in fixed costs), then markups are likely to differ.¹⁵

¹⁴ As in Berndt and Fuss (1986), the appropriate input “price” is the factor’s shadow rental cost.

¹⁵ An earlier version of this paper derived the algebra in terms of returns to scale rather than markups. The two approaches emphasize different aspects of the same issues. The markup-approach helps focus on technology versus productivity; the returns-to-scale approach clarifies how aggregation affects estimates of returns to scale.

Following Hall (1990), cost minimization¹⁶ implies that output growth, dy_i , equals the markup multiplied by revenue-share-weighted input growth, dx_i , plus gross-output-augmenting technology change, $\frac{F_T^i T_i}{F^i} dt_i$. So for any input J and firm i , where s_{Ji} is the input's share in nominal gross output and dJ is the input's growth rate,

$$(15) \quad \begin{aligned} dy_i &= \mu_i [s_{Li} dl_i + s_{Ki} dk_i + s_{Mi} dm_i] + \frac{F_T^i T_i}{F^i} dt_i \\ &\equiv \mu_i dx_i + \frac{F_T^i T_i}{F^i} dt_i. \end{aligned}$$

As in Sections I and II, the revenue shares need not sum to one if there are economic profits or losses. However, if there are no economic profits, then our revenue shares coincide with Solow's shares.

Since our ultimate interest is in value added, we now derive the value-added analogues to (15).

From the production side, we use the standard Divisia definition of firm-level value-added, dv_i :

$$(16) \quad dv_i = \frac{dy_i - s_{Mi} dm_i}{1 - s_{Mi}} = dy_i - \left[\frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i).$$

With some algebraic manipulation (shown in Appendix A), we can write dv_i as:

$$(17) \quad dv_i = \left[\frac{\mu_i (1 - s_{Mi})}{1 - \mu_i s_{Mi}} \right] dx_i^V + \left[\frac{\mu_i s_{Mi}}{1 - \mu_i s_{Mi}} - \frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i) + \frac{F_T^i T_i}{F^i} \frac{dt_i}{1 - \mu_i s_{Mi}},$$

where primary-input growth, dx_i^V , is defined analogously to aggregate primary input growth:

$$(18) \quad dx_i^V = \frac{s_{Ki}}{1 - s_{Mi}} dk_i + \frac{s_{Li}}{1 - s_{Mi}} dl_i \equiv s_{Ki}^V dk_i + s_{Li}^V dl_i.$$

¹⁶ Contrary to some of the statements in the literature, the derivation does not require profit-maximization. Hence the relationship we derive below is robust to any form of price-setting behavior; for example, it allows for sticky output prices and for complex dynamic pricing strategies derived from supergames (e.g. Rotemberg and Saloner [1986]). See Basu and Fernald (1995, Appendix). In particular, note that markups need not depend just on the elasticity of demand.

Equation (17) looks complicated, but some further assumptions on the production function allow us to simplify it. Suppose that the production function in (1) takes the following separable form:

$$(19) \quad Y_i = F^i(K_i, L_i, M_i, T_i) = G^i(V^{Pi}(K_i, L_i, T_i), H^i(M_i)),$$

The firm combines primary inputs to produce “productive value added”, V^{Pi} , which it then combines with intermediate inputs to produce gross output. We can break the cost-minimization problem into two stages: first, minimize the cost of using primary inputs to produce any level of V^{Pi} ; second, minimize the cost of using productive value added and intermediate inputs to produce any level of gross output.

In the first stage, the logic from equation (15) implies that the “productive” value-added growth, dv^P , depends on the revenue-weighted growth in primary inputs dx^V , plus technology shocks (without loss of generality we normalize to one the elasticity of productive value added V^{VV} with respect to technology):

$$(20) \quad dv_i^P = \mu_i^V dx_i^V + dt_i.$$

The value-added markup μ_i^V equals the ratio of the price of productive value added to the marginal cost of producing it. Appendix A shows that under standard conditions:

$$(A5) \quad \mu_i^V = \mu_i \frac{1 - s_{Mi}}{1 - \mu_i s_{Mi}}.$$

Returning to equation (17), we can now rewrite it as follows:

$$(21) \quad dv_i = \mu_i^V dx_i^V + (\mu_i^V - 1) \left[\frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i) + dt_i.$$

Real value-added growth depends on primary input growth, changes in the materials-to-output ratio, and technology. The first term shows that primary inputs are multiplied by the value-added markup. The second term reflects the extent to which the standard measure of value added differs from “productive” value added V^{Pi} , and hence does not properly measure the productive contribution of intermediate inputs. Intuitively, the standard measure of value added subtracts off intermediate input growth using

revenue shares, whereas with imperfect competition the productive contribution of these inputs exceeds the revenue share by the markup. The third term is the value-added-augmenting technology shock.

The firm's revenue-weighted value-added productivity residual, dp_i , equals $dv_i - dx_i^V$. Hence,

$$(22) \quad dp_i = (\mu_i^V - 1)dx_i^V + (\mu_i^V - 1) \left[\frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i) + dt_i.$$

Firm-level productivity growth measured in terms of value added depends in part on markups, as emphasized by Hall. In the presence of imperfect competition, however, productivity growth also depends positively on changes in the relative intensity of intermediate-input use.

B. Aggregating over Firms

We now aggregate over firms to the economy-wide level. Technology shocks clearly affect measured aggregate productivity. In addition, aggregate productivity depends on changes in aggregate primary inputs, changes in the distribution of inputs across firms (when inputs have different marginal products in different uses), and changes in the intensity of intermediate input use.

Aggregate productivity growth is the difference between the growth rates of aggregate output, dv , and aggregate inputs, dx^V . From the production side of the national accounts identity, aggregate output is a Divisia index of firm-level value-added. In growth rates:

$$(23) \quad dv \equiv \sum_{i=1}^N w_i dv_i,$$

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

$$w_i \equiv \frac{P_i^V V_i}{P^V V}.$$

With some algebraic manipulation, we can write aggregate primary-input growth dx^V in terms of the weighted average of firm-level input growth, and reallocations of capital and labor:

$$(24) \quad dx^V = \sum_{i=1}^N w_i dx_i^V - R_K - R_L,$$

where

$$R_K = \sum_{i=1}^N w_i s_{Ki}^V \left[\frac{P_{Ki} - P_K}{P_{Ki}} \right] dk_i,$$

$$R_L = \sum_{i=1}^N w_i s_{Li}^V \left[\frac{P_{Li} - P_L}{P_{Li}} \right] dl_i.$$

Combining equations (23) and (24), above, and noting the definition of firm-level productivity (equation (22)), we can write aggregate productivity as:

$$(25) \quad dp = \sum_{i=1}^N dp_i + R_K + R_L.$$

As we have defined the input prices, they represent differences in *shadow values* across uses (which may or may not be reflected in factor-price differences). Thus, aggregate productivity is the weighted average of firm-level productivity shocks, plus reallocations of capital and labor among uses with different shadow values. If resources shift towards more highly valued uses, then aggregate productivity can rise even with no change in firm-level productivity. Apart from input-reallocations, this equation shows that any deviations between aggregate productivity and aggregate technology must be coming from differences between firm-level productivity and firm-level technology.

By substituting from equation (22) for dp_i and manipulating, we can write aggregate productivity in terms of technology, aggregate inputs, and reallocations of resources:

$$(26) \quad dp = (\bar{\mu}^V - 1)dx^V + R_\mu + R_M + \bar{\mu}^V R_K + \bar{\mu}^V R_L + dt$$

where,

$$\bar{\mu}^V = \sum_{i=1}^N w_i \mu_i^V,$$

$$R_\mu = \sum_{i=1}^N w_i (\mu_i^V - \bar{\mu}^V) dx_i^V,$$

$$R_M = \sum_{i=1}^N w_i (\mu_i^V - 1) \left[\frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i), \text{ and}$$

$$dt = \sum_{i=1}^N w_i dt_i.$$

We define the sum of the reallocation terms as R , where

$$(27) \quad R = R_{\mu} + R_M + \bar{\mu}^V R + \bar{\mu}^V R .$$

Equation (26) shows the distinction between aggregate productivity and aggregate technology. If every firm is perfectly competitive and pays the same price for factors (perhaps because factors are completely mobile), then all terms other than dt disappear: productivity equals technology. However, with imperfect competition or frictions in product or factor markets, productivity and technology are not equivalent.¹⁷

C. The Definition of Technology Change

Conceptually, aggregate technology change should measure the change in aggregate output in response to firm-level technology shocks, when primary inputs remain fixed. Under what conditions does this correspond to our measure dt ?

Suppose all markets are perfectly competitive. Domar (1961) and Hulten (1978) discuss this case, and find that our definition properly measures the outward shift in society's PPF in response to firm-level technology shocks. In Figure 1 again, this case corresponds to a point like A , the optimal production point, where society's resources are optimally allocated.

However, consider points B and C in Figure 1. Aggregate technology is not defined unambiguously at these points, since different changes in aggregate output are consistent with the same firm-level technology shocks. In particular, technology shocks may affect the distribution of inputs across firms. For example, if technology shocks cause a reallocation of inputs from low- to high-markup firms, aggregate output rises more than if there were no reallocations. Even with a fixed distribution of

¹⁷ Jorgenson, Gollop and Fraumeni (1987) derive an equation for the case of constant returns to scale and perfect competition, so that $dp_i = dt_i$. Thus, they omit the terms other than R_K and R_L . They also allow for heterogeneity in capital and labor, which we have ignored for simplicity. With heterogeneity, our results generalize easily: For example, if R_{Kk} is the factor-price reallocation term for capital of type k , then $R_K = \sum_k R_{Kk}$.

primary inputs, technology shocks might affect the intensity of intermediate-input use, thereby changing output growth.

Given this ambiguity, we define aggregate technology so that it correctly measures the increase in aggregate output under standard conditions where an aggregate production function exists.¹⁸ Suppose all firms have the same separable gross-output production function as in equation (19), and always use materials in fixed proportions to gross output (so that $dm = dy$ always). Suppose also that there are no factor-market frictions (so that R_K and R_L are identically zero) and that all firms charge the same markup in output markets. Under these assumptions, which are implicit or explicit in most dynamic general-equilibrium models with imperfect competition, there is a representative producer and an aggregate value-added production function. In this case, aggregate technology change corresponds to our definition dt .

D. Productivity Interpretation

The markup-reallocation term, R_μ , represents shifts of resources toward uses with higher social valuations. The markup represents the wedge between the social valuation of a good and its marginal cost. Reallocating resources from low- to high-markup firms thus shifts resources towards uses where consumers value them more highly. Aggregate output combines goods using consumers' valuations of them, so aggregate output and productivity rise.

If the variability of firms' inputs are correlated with their market power, then imperfect competition affects aggregate productivity even if the average markup is small. For example, Basu and Fernald (1997) estimate that durable-goods industries have larger returns to scale and markups than non-durable-goods industries. Durable-industries are more cyclical, and employ a larger share of the

¹⁸ The discussion in this paragraph follows Rotemberg and Woodford (1995).

marginal inputs in a boom. This marginal reallocation thus contributes to the procyclicality of aggregate productivity.

The input-reallocation terms R_L and R_L also reflect reallocations of resources among uses with different social valuations. Consider labor. Shifting labor from firms where it has a low shadow value to firms where it has a high shadow value increases aggregate output. Why might shadow values (or wages) differ across firms? First, labor may not be instantaneously mobile across sectors; sectoral shifts may lead workers in, say, defense industries to have lower marginal products than they would in health-care.¹⁹ Second, efficiency wage considerations may be more important in some industries than others, as emphasized by Katz and Summers (1989). Third, unions with monopoly power might choose to charge different wages to different firms. Whatever the reason, shifting labor to more productive uses increases aggregate output, even if total input does not change.

Note that the first reason, costly factor mobility, is completely consistent with constant returns and perfect competition. Differences in marginal products that reflect factor immobility should be temporary. Although we did not model adjustment costs explicitly, the literature on these costs shows that the appropriate “factor prices” are the shadow values of the factors to the firm.²⁰ With costly factor adjustment, these shadow values may differ substantially; hence reallocation effects on output and productivity may be significant, even in a world with perfect competition and constant returns.²¹

The materials-reallocation term, R_M , reflects the extent to which measured real value added depends on the intensity of intermediate-input use. Firm-level value added is useful for national

¹⁹ Whether differences in labor’s marginal product lead to differences in wages depends on whether the adjustment costs are paid by workers or firms.

²⁰ See Berndt and Fuss (1986).

²¹ The “sectoral shifts” literature takes this approach; see, e.g. Phelan and Trejos (1996). Horvath (1995) also incorporates adjustment costs into a dynamic general equilibrium model, generating effects on aggregate productivity from input reallocations. Microeconomic productivity literature (e.g. Baily et al. (1992)) finds that there are systematic productivity differences across firms within narrowly defined industries; these productivity differences show up either as higher profits from a higher markup, and hence are reflected in R_M , or higher factor payments, and hence are reflected in R_K and R_L .

accounting, regardless of technology or market structure. However, it does not in general correspond to “productive” value added V^{Pi} , defined in equation (19). In particular, with imperfect competition, the construction of value added does not subtract off the full marginal product of intermediate inputs, since the marginal product of these goods exceeds their cost; there is a wedge between each firm’s measured real value added and productive value added. R_M , which equals the sum of these wedges, represents real goods and services, and hence affects aggregate output and productivity.

Note that R_M depends on the size of markups in firms *using* materials. Consider an economy where some firms produce intermediate goods using capital and labor, and other firms assemble intermediate goods into final goods (e.g. Beaudry and Devereux (1994)): the importance of R_M depends on the size (and heterogeneity) of markups in the *final-goods* industry. This is important because firms may be able to negotiate multi-part prices with long-term suppliers of their inputs, and thus partially offset the inefficiencies resulting from imperfect competition in intermediate-goods industries.²² The inefficiency that we identify in R_M is, however, a function of the markups of the firms *using* intermediate goods, not those *selling* such goods; multi-part pricing for intermediate goods does not eliminate this inefficiency. (However, the inefficiency is larger in symmetric models where all output is also used as materials input: we discuss this “double marginalization” effect in the next subsection.)

E. Welfare Interpretation

In Section II we argued that productivity measures welfare. We now provide intuition for why each term in equation (26) represent a welfare improvement.

The welfare effect of an average markup greater than one reflects the fact that markups distort the labor-leisure choice. Consumers would prefer to provide more capital and labor and consume the extra

²² We thank Robert Hall for this observation.

market goods they could produce, since the utility value of these goods exceeds the disutility of producing them (through providing capital and labor). Hence welfare rises when inputs, dx^V , rise.

This intuition also encompasses R_μ . Suppose a firm has a higher-than-average markup. Then relative to the social optimum, that firm produces even less output than does the average firm. Thus, if inputs rise in that firms, welfare increases even more than if input use had risen in the average firm.

The intuition for the materials term R_M comes in two parts. First, there is the standard effect that markups reduce input use, as discussed in our analysis of the average-markup term and R_μ . This effect depends, of course, on the size of the markup in the firm using materials. If the production function requires that firms use materials in fixed proportions to gross output, then the suboptimally low usage of primary inputs (capital and labor) is a sufficient statistic for the suboptimally low usage of materials. In this case, dm identically equals dy , and the materials term (which depends on $dm - dy$) drops out.

Second, if firms can substitute away from materials in producing gross output — and the best evidence argues that they can²³ — then there is an *independent* distortion from using too few materials. Markups on materials reduce productive efficiency, leading the economy to produce within its production possibilities frontier rather than simply at an inefficient point on the frontier.²⁴ Basu (1995) presents a model where materials need not be used in fixed proportions with output, and finds that with markup pricing firms generally use too few materials relative to the quantity of primary inputs they purchase: thus, materials intensity is too low. In his model, $dm - dy$ is procyclical because countercyclical markups increase materials intensity in booms. Increasing materials intensity increases welfare by moving the economy closer to the boundary of the PPF. Rotemberg and Woodford (1996) find that including this effect in an imperfectly-competitive dynamic general equilibrium model

²³ See Bruno (1984) and Rotemberg and Woodford (1996, Appendix).

²⁴ This point is similar to that of Diamond and Mirrlees (1971), who show that taxes on intermediate goods generally distort relative factor proportions and move the economy within its PPF. From the point of view of the materials-using firm, markups are equivalent to a tax on intermediate goods. See also Judd (1995).

significantly improves the model’s ability to match the dynamics of output and the real wage following an exogenous increase in oil prices.

The welfare interpretation of R and R depends on their source. Our proof in Section II does not apply directly to these terms, since we assumed that capital and labor had unique factor prices. The key to our proof is that the wage equal the marginal disutility of labor. The analysis in Section II easily extends to the case where wages, say, differ across firms because of costly labor mobility; productivity change still corresponds to welfare change. On the other hand, suppose workers are rationed in the amount of labor they can supply to certain firms (perhaps because of efficiency wages in some firms but not others). In the firms where labor is rationed, the wage does not equal the marginal disutility of labor, so the proof in Section II does not apply. (With rationing, aggregate labor input must be weighted by the marginal wage — the wage in the firm(s) where labor input is not rationed — rather than the average wage.) We plan to explore these issues in future work, but we do not pursue them further in this paper.

Finally, technology improvements *ceteris paribus* obviously raise welfare, since consumers get more output from given inputs. However, in general equilibrium with distortions, welfare could actually *fall*. For example, technology improvements could lead sectors with smaller-than-average markups to increase input use more than average, making the R_μ term negative.²⁵

IV. Data and Method

A. Data

Having discussed the theory of aggregation, we now explore its empirical significance. We construct a measure of “true” aggregate technology change, dt , and compare it to aggregate productivity

²⁵ Bai et al. (1997) make essentially this argument for state enterprises in China. They argue that state firms set prices *below* marginal cost in order to maximize sales; technology improvements in that sector then lead it to increase input use and thus reduce national welfare.

growth, dp . We estimate technology change at a disaggregated level, and then aggregate. Our aggregate is the private U.S. economy, and our “firms” are 34 industries at roughly the two-digit SIC level.

The average two-digit manufacturing industry has about 18,000 enterprises, so it may seem odd to take industries as firms. Unfortunately, there are no firm-level data sets that 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 — necessary for the welfare analysis in Section II — as well as panel length and data quality. By focusing on aggregates, our work complements existing work that examines small subsets of the economy. Nevertheless, the aggregation effects we identified in Section III are probably important even within the disaggregated sectors that we use. This raises several concerns that we discuss in the next subsection and in Appendix B.

We use data compiled by Dale Jorgenson and Barbara Fraumeni on industry-level inputs and outputs. These data consist of a panel of 34 private industries (including 21 manufacturing industries) that constitute the entire U.S. private business economy for the years 1959-1989.²⁶ 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 our purposes, an essential aspect of the data is their inclusion of intermediate inputs.²⁷

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), Hall (1990), and Caballero and Lyons (1992), and estimate the user cost of capital r . For any type of capital, the required payment is then rP_KK , where P_KK is the current-dollar value of the stock of this type of capital. In each sector, we use data on the

²⁶ We restrict our sample to the period 1959-1989 because for those years the output data in the Jorgenson dataset come from a consistent source, the Bureau of Labor Statistics.

²⁷ For a complete description of the dataset, see Jorgenson et al. (1987).

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, the user cost of capital is

$$r_s = (\rho + \delta_s) \frac{(1 - ITC_s - \tau d_s)}{(1 - \tau)}, \quad s = 1 \text{ to } 53.$$

ρ 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 ρ 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 the cost shares is straightforward.

For our empirical work, we need instruments that are uncorrelated with technology change. We use versions of the Hall-Ramey instruments: the growth rate of the price of oil deflated by the GDP deflator, the growth rate of real government defense spending, and the political party of the President.

B. Estimating Technology Change

In order to estimate “firm-level” technology change, we treat equation (15) as an estimating equation for the gross-output markup, μ . The sum of the constant and the residual then measures technology change in the gross-output production function. Our derivations present an accounting identity for output change at each moment in time; their discrete-time empirical counterparts (e.g. equation (15), implemented as a regression) are best thought of as first-order approximations.

In order to avoid the “transmission problem” of correlation between technology shocks and input use, we use the Hall-Ramey instruments noted above. We use the current value and one lag of each instrument. (We experimented with subsets of the instrument list — e.g. using only contemporaneous values of the instruments. Our results were not too sensitive to the choice, so we report results only for the largest set.) For reasons discussed in Appendix B, we prefer the instrumented results. However, we

report OLS results for comparison, and because of questions about the small-sample properties of the Hall-Ramey instruments (see, for example, Burnside (1996)).

Since we are ultimately interested in aggregate final expenditure, we transform the gross-output technology shocks into value-added technology shocks. The relationship between the two is given by equation (A7). We use our estimated markups to do the transformation. We then aggregate these shocks using value-added weights. The difference between aggregate productivity and aggregate technology gives us the sum of our reallocation terms, R , as defined in equation (26). We are also interested in the components of R . We estimate dt from the sectoral regressions; with the estimates of μ from those regressions we can also construct R_μ and R_M . Since we also know dp , equation (26) shows that we can estimate of $\bar{\mu}^V R_K + \bar{\mu}^V R_L$, but we cannot construct each term separately.

Empirically, one concern is that some of the sectoral estimates of μ are less than one; it is *a priori* implausible that these measure true firm-level markups. These estimates presumably reflect aggregation from the firm level to the industry level, where the unobserved within-industry reallocations are negatively correlated with industry inputs. In Appendix B, we discuss the implications of these reallocations for our results. We argue that our procedure of using industry-level data does not create any necessary bias in our results, and may be preferable to using small firm-level data sets.

V. Results

We now investigate the relationship between aggregate productivity growth and aggregate technology growth. We find that technology has about the same mean growth rate as productivity, but is less volatile and less correlated with both input and output growth than is aggregate productivity growth. As we discuss, these results are important for judging the plausibility of standard RBC models.

We define aggregate productivity growth as the welfare-relevant modified Solow residual defined in equation (3). This measure differs from the standard Solow residual since the weights do not

necessarily sum to one; the difference reflects economic profits or losses. However, since estimated profits are small (about 3 percent on average), the results we report are essentially unchanged if we use the standard Solow residual instead. Aggregate technology change is the weighted sum of sectoral value-added technology changes, as described in Section V. We report results with two technology measures, one using residuals from instrumented regressions and the other using residuals from uninstrumented regressions.

Table 1 presents summary statistics for private output growth, dv ; the weighted average of primary input growth, dx^V ; the modified Solow residual measuring productivity growth; and the two technology series. Table 2 is a correlation matrix for these variables. All results are for the sample period 1959-89.

Table 1 shows that at the aggregate level our two technology series have significantly lower variance than does productivity growth: about 60 percent of the variance of dp . However, we estimate that they have approximately the same average growth rate (and one cannot reject the hypothesis that the means are equal). This result confirms our intuition that the reallocation effects are indeed most relevant at business-cycle frequencies, instead of being growth phenomena. We return to this point below.

The correlations in Table 2 show striking differences between productivity growth and technology change. The three series are fairly highly correlated, with correlation coefficients of 0.88 and 0.81. But the correlation of productivity growth with output growth is 0.85, whereas the OLS measure of technology change has an output correlation of 0.67, and the IV measure only 0.53. And while productivity growth has a positive correlation with aggregate inputs of 0.23, the OLS measure of technology growth shows little correlation — 0.05 — and the IV measure has a *negative* correlation: -0.13. The correlations with hours worked give a similar picture: productivity growth has a correlation of 0.43, our OLS technology measure has a correlation of 0.13, and the IV technology measure has a correlation of -0.12.

We present summary statistics for the components of the reallocation term in Table 3. Examining the components of the reallocation term, the ones that seem to be most important are R_μ and R_M . For

the instrumented series, these terms have standard deviations of 0.72 and 0.51 respectively, out of a total standard deviation of 1.11 for R . The sum of R and R (we are unable to separate out the two), is much less volatile.

One might ask whether the reallocation effects we have identified represent the important gap between productivity and technology, or whether the difference between the two is still driven mostly by the “average markup” effect identified by Hall (1988, 1990). We compared our results to what we would have found had we used only the “average” correction, i.e. that coming from the presence of the $(\bar{\mu}^V - 1)dx^V$ term. Using the correlations with dx^V as a benchmark, we find that the reallocation effects are the more important. For the OLS results, productivity corrected for the average effect would have yielded a correlation with input growth of 0.18, as opposed to 0.23 for the uncorrected series and 0.05 for our estimated technology series. For the IV results, the correlation would have been 0.12, as opposed to -0.13 for the estimated series. Thus, our reallocation effects are responsible for at least two-thirds of the correction. This result should not be surprising since, as we noted, the recent literature finds small average markups. Our results echo this finding; for the OLS case, we find $\bar{\mu}^V = 1.05$, and for the IV case $\bar{\mu}^V = 1.13$. The surprising result is that even such small average markups are consistent with important differences between aggregate productivity and aggregate technology coming from reallocations across sectors.

Altogether, these findings indicate that standard technology-driven RBC models are even further from fitting the data than one normally thinks. These models in any case tend to generate a correlation between output and technology that is much higher than the correlation in the data between output and the Solow residual; since our estimates reduce the measured correlation between output and technology by 30 percent or so, then standard RBC models are that much further from the truth. More importantly, business-cycle models must display the standard characteristic of business cycles: a comovement

between output and inputs. If inputs and output move independently in response to technology shocks, then technology shocks cannot be the dominant impulse driving business cycles.²⁸

However, it is possible that technology shocks contribute in a different way to business cycles, since composition effects themselves may serve as a major propagation mechanism for technology shocks.²⁹ For example, suppose a technology improvement (contemporaneous or lagged) boosts output by more in durable- than in non-durable-goods sectors. Since Basu and Fernald (1997) find that durable-goods industries have larger markups, such a reallocation increases output and productivity. Hence, though we find that output is less correlated with *contemporaneous* technology shocks after composition corrections, it is possible that *lagged* technology shocks themselves drive composition changes. It is thus possible for our results to support technology-driven models by reducing the size of the necessary technology shocks without reducing their ability to explain output fluctuations.

To investigate this possibility, we regress the reallocation component of productivity growth, R , on lags of our derived series of technology shocks. Table 4 shows these results, for both the OLS and IV estimates. Composition changes do appear to help propagate lagged technology shocks — the results are both statistically and economically significant. However, the sign of the effect is *negative*: lagged technology improvements tend to worsen the allocation of inputs in the economy. This conclusion is not *prima facie* absurd. For example, technology improvements may stimulate demand that cannot be satisfied by the output of efficient firms. As these firms hit capacity constraints, inefficient firms may produce a larger share of output. Then the reallocation effect would be negative, exactly as we find. Thus, the reallocation effects we identify do not serve to amplify the effects of lagged technology

²⁸ As we noted, some of the sectoral regressions have estimated values of “ μ ” that are smaller than one. In Appendix B, we argue that this is not likely to be a problem for the main results. Nevertheless, we do try restricting the values to be equal to or greater than one, by setting all point estimates that are less than one to be equal to one. This has a relatively small effect on the correlations: for the instrumented estimates, the correlation of dt_i^V with aggregate output becomes 0.59, while the correlation with aggregate inputs remains -0.13.

²⁹ We thank Martin Eichenbaum and Michael Horvath for suggesting this possibility.

improvements —the reverse is true. (We do find that lagged technology improvements significantly increase aggregate output and inputs, which is what one would expect.)

Finally, we note that our results do not shed light on the celebrated puzzle of the productivity slowdown in the early 1970s. Of course, reallocations cannot be a source of long-term productivity growth (since their maximum size is confined to a limited range). However, over any finite sample period composition effects can account for a significant mismeasurement in the average technology growth rate. Our results do not support this view, however. In Table 5, we report the summary statistics of Table 1 for productivity growth and our two technology series, for the time periods 1959-73 and 1974-89. All three series show essentially the same reduction in the mean growth rate. (Interestingly, all three series also show significantly higher volatility after 1973).

Figure 2 plots productivity and the IV technology series. Both are normalized to equal 1 in 1959, with the changes cumulated thereafter. Both grow rapidly until the oil price shock of 1974, with much more modest growth thereafter. However, productivity falls much more in 1974 and in the 1979-82 period, thus accounting for the higher correlation with the business cycle.

Our results indicate that the productivity slowdown was indeed a slowdown in the rate of technical progress, not a series of shocks that worsened the economy's distribution of inputs. This conclusion matters for policy: appropriate policies, if any, are those that spur innovation — e.g. funding for basic and applied research — not various types of industrial policy designed to protect “good” (i.e., high-rent) jobs.

VI. Conclusion

In this paper, we explore the meaning of aggregate productivity in a world with frictions and distortions. In such a world, productivity growth does not in general estimate aggregate technology change. Nevertheless, a modified Solow residual does measure welfare change. Thus, though much of

the recent productivity literature emphasizes the use of micro data, welfare measurement requires only readily-available national income accounts data.³⁰

Of course, technology shocks are an important source of productivity change. We provide a general accounting framework that relates growth in aggregate productivity and aggregate technology. We identify various non-technological terms that reflect changes in the allocation of factors across uses with different marginal products. Marginal products, in turn, can differ because of frictions or distortions in the economy. These reallocations affect aggregate output and productivity, without necessarily reflecting technology. Hence, computing aggregate technology change requires micro data.

Several existing studies provide models of fluctuations in economies that deviate in a variety of ways from the standard one-sector model of production. In those models, the non-technological sources of productivity fluctuations are not always clear, nor is the relationship to other models. Our general framework can aid in understanding and interpreting the fluctuations arising, for example, from sector-specific technology shocks, vintage capital effects, or imperfect competition with heterogeneity (see, respectively, Phelan and Trejos (1996), Gilchrist and Williams (1996), and Basu et al. (1996)).

Our derivation also suggests a new technique for measuring an economy's allocative efficiency. Suppose we compute technology change from disaggregated data and productivity growth from aggregate data. The difference between the two represents changes in allocative efficiency. In principle, this technique could be used to calculate the effects of a policy change that improves the distribution of existing resources; e.g., the post-socialist transition in Eastern Europe.

Finally, we apply our decomposition to the data. Composition changes appear to explain much of the cyclical of aggregate productivity. We estimate aggregate technology shocks by aggregating industry-level shocks. Compared with the productivity residual, technology shocks are less volatile,

³⁰ Of course, we have assumed that national income comprises *all* of an economy's output. In practice, there are formidable difficulties in measuring some of these outputs: e.g. household production, investment in human capital, and changes in environmental quality.

have a significantly smaller contemporaneous correlation with output, and have basically zero contemporaneous correlation with inputs. These differences are important for judging the plausibility of technology-driven RBC models, with or without imperfect competition.

Composition changes also appear to serve as a significant propagation mechanisms for lagged technology shocks. Surprisingly, lagged technology shocks appear to have a *negative* effect on the subsequent allocation of inputs. Taken as a whole, our results do not support the standard technology-driven RBC model, where technology must be correlated with inputs.

Lagged technology shocks account for only about 20 percent of the variation in the composition effects, which in turn contribute significantly to the variability of output and productivity. Thus, composition effects are an important economic mechanisms in their own right, and may contribute to the amplification and (positive) propagation of impulses other than technology shocks. Basu et al. (1996) show that induced reallocations significantly propagate government spending shocks in a multi-sector dynamic general-equilibrium model with imperfect competition. This avenue of research may help address the notorious weakness of the propagation mechanisms in the one-sector RBC model (see, e.g., Cogley and Nason (1995)).

Future research should refine our estimates of aggregate technology change (and hence of the cyclicity of composition effects). Because of our interest in economy-wide aggregates, our “disaggregated” data is at the level of two-digit industries. In principle, our exercise could be repeated for the manufacturing sector with a large firm-level data set, such as the Census’s Longitudinal Research Database. In addition, we abstracted from cyclical measurement error, which also causes measured productivity growth to differ from technology growth. Cyclical mismeasurement is conceptually very different from the issues addressed in this paper, since it does not lead to the same welfare conclusions for aggregate productivity. However, such measurement error, most likely coming from unobserved

variations in capital and labor utilization, may be quite significant.³¹ Thus, ideally one would adjust firm-level productivity for variations in utilization to get a true firm-level technology series, and then aggregate technology change across firms using our methods. These and similar projects are likely to shed new light on the impulses and propagation mechanisms driving business cycles.

³¹ See, for example, Burnside et al. (1995), Basu (1996) and Shapiro (1996). Griliches (1987) discusses many of the measurement issues involved in calculating productivity.

Appendix A: Derivations in Section III

This appendix derives the equations in Section III in greater detail. From the production function (11), we can write output growth dy in terms of input growth and output elasticities:

$$(A1) \quad dy_i = \frac{F_L^i L_i}{F^i} dl_i + \frac{F_K^i K_i}{F^i} dk_i + \frac{F_M^i M_i}{F^i} dm_i + \frac{F_T^i T_i}{F^i} dt_i.$$

From the first-order conditions (13), the output elasticities are related to factor shares and markups. That is, for any input J :

$$(A2) \quad \frac{F_J^i J_i}{F^i} = \mu_i s_{Ji}.$$

From the two equations above, equation (15) in the text for gross output growth follows:

$$(A3) \quad \begin{aligned} dy_i &= \mu_i [s_{Li} dl_i + s_{Ki} dk_i + s_{Mi} dm_i] + \frac{F_T^i T_i}{F^i} dt_i \\ &\equiv \mu_i dx_i + \frac{F_T^i T_i}{F^i} dt_i. \end{aligned}$$

Our goal is to express the growth rate of value added, dv , in terms of primary inputs and the intensity of intermediate-input use. To facilitate this derivation, note first that we can write the above equation as

$$\begin{aligned} dy_i &= \mu_i (1 - s_{Mi}) \left[\frac{s_{Li}}{(1 - s_{Mi})} dl_i + \frac{s_{Ki}}{(1 - s_{Mi})} dk_i \right] + \mu_i s_{Mi} dm_i + \frac{F_T^i T_i}{F^i} dt_i \\ &= \mu_i (1 - s_{Mi}) [s_{Li}^V dl_i + s_{Ki}^V dk_i] + \mu_i s_{Mi} dm_i + \frac{F_T^i T_i}{F^i} dt_i \\ &= \mu_i (1 - s_{Mi}) dx_i^V + \mu_i s_{Mi} dm_i + \frac{F_T^i T_i}{F^i} dt_i. \end{aligned}$$

This expression relates gross output growth to growth in primary inputs, dx_i^V , intermediate inputs, dm_i , and technology shocks. Primary input growth is defined as a weighted average of capital and labor growth, using shares in value added.

Second, subtract $\mu_i s_{Mi} dy_i$ from both sides, and divide through by $(1 - \mu_i s_{Mi})$. Output growth can then be written as:

$$dy_i = \left[\frac{\mu_i (1 - s_{Mi})}{1 - \mu_i s_{Mi}} \right] dx_i^V + \left[\frac{\mu_i s_{Mi}}{1 - \mu_i s_{Mi}} \right] (dm_i - dy_i) + \frac{F_T^i T_i}{F^i} \frac{dt_i}{1 - \mu_i s_{Mi}}.$$

Given the Divisia definition of value-added growth dv (equation (16) in the text), equation (17) in the text follows immediately:

$$dv_i = \left[\frac{\mu_i (1 - s_{Mi})}{1 - \mu_i s_{Mi}} \right] dx_i^V + \left[\frac{\mu_i s_{Mi}}{1 - \mu_i s_{Mi}} - \frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i) + \frac{F_T^i T_i}{F^i} \frac{dt_i}{1 - \mu_i s_{Mi}},$$

It is obvious from this equation that value-added growth is not, in general, simply a function of primary inputs dx_i^V . A long literature in the 1970s explored whether a “value added function” exists, and argued that the answer depended on separability properties of the production function. The equation above shows that with imperfect competition, taking value-added to be a function only of primary inputs is generally misspecified — regardless of whether the production function is separable between value added and intermediate inputs.

Nevertheless, it will be useful to make that further assumption of separability in order to provide a simple interpretation of value-added growth. In particular, suppose the production function is separable, as in equation (19):

$$Y_i = F^i(K_i, L_i, M_i, T_i) = G^i(V^{Pi}(K_i, L_i, T_i), H^i(M_i)),$$

The producer combines primary inputs to produce “productive value added”, V^{Pi} , then combines productive value added with intermediate inputs to produce gross output. We can break the cost-minimization problem into two stages: first, minimize the cost of using primary inputs to produce any level of V^{Pi} ; second, minimize the cost of using productive value added and intermediate inputs to produce any level of gross output (taking the marginal cost of producing V^{Pi} , from the first stage, as the transfer price of productive value added in the second stage).

In the first stage, the logic used to derive equation (15) implies that the growth rate of “productive” value added, dv^P , can be written in terms of the revenue-weighted growth in primary inputs dx^V , plus technology shocks (without loss of generality we normalize to one the elasticity of productive value added V^P with respect to technology):

$$dv_i^P = \mu_i^V dx_i^V + dt_i.$$

In the second stage, the first-order condition for value-added implies that:

$$(A4) \quad \frac{G_V V^{P_i}}{G} = \frac{\mu}{\mu^V} s_{V_i}.$$

Note that s_{V_i} equals $P_i^V V_i / P_i Y_i = (P_i Y - P_{M_i} M_i) / P_i Y$, which equals $(1 - s_{M_i})$. However, without knowing more about the shape of the production function (and hence, the slopes of marginal cost of producing V^P and Y), we cannot make any general statements about the magnitude of the value-added markup μ_i^V .

To do so, we make the further substantive assumption that all returns to scale are in V^P , arising perhaps from overhead capital or labor. This requires that G be homogeneous of degree one in V^P and H , and that H be homogeneous of degree one in M . Under these assumptions, the left-hand-side of equation (A4) equals $(1 - \mu_i s_{M_i})$. We can thus rearrange the first-order for productive value added to show:

$$(A5) \quad \mu_i^V = \mu_i \frac{1 - s_i}{1 - \mu_i s_{M_i}}.$$

Note also that

$$(A6) \quad \frac{F_T^i T}{F^i} dt_i = (1 - \mu_i s_{M_i}) dt_i.$$

Returning to equation (17), we can now rewrite it as in the text:

$$dv_i = \mu_i^V dx_i^V + (\mu_i^V - 1) \left[\frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i) + dt_i.$$

Equation (22) for firm-level value-added productivity follows immediately.

We now aggregate over firms to get aggregate productivity growth, which is the difference between the growth rates of aggregate output and aggregate primary inputs. We defined aggregate primary input growth as $dx^V = s_K^V dk + s_L^V dl$. Using the definitions of s_{Ki}^V and s_{Li}^V , and differentiating the definitions of aggregate K and L, we can write this as:

$$\begin{aligned} dx^V &= \frac{P_K K}{P^V V} \sum_{i=1}^N \frac{K_i}{K} dk_i + \frac{P_L L}{P^V V} \sum_{i=1}^N \frac{L_i}{L} dl_i \\ (A7) \quad &= \sum_{i=1}^N \frac{P_i^V V_i}{P^V V} \frac{P_{Ki} K_i}{P_i^V V_i} \frac{P_K}{P_{Ki}} dk_i + \sum_{i=1}^N \frac{P_i^V V_i}{P^V V} \frac{P_{Li} L_i}{P_i^V V_i} \frac{P_L}{P_{Li}} dl_i. \end{aligned}$$

Noting the definitions of w_i , s_{Ki}^V , and s_{Li}^V , we can write this as:

$$\begin{aligned} dx^V &= \sum_{i=1}^N w_i s_{Ki}^V \frac{P_K}{P_{Ki}} dk_i + \sum_{i=1}^N w_i s_{Li}^V \frac{P_L}{P_{Li}} dl_i \\ (A8) \quad &= \sum_{i=1}^N w_i (s_{Ki}^V dk_i + s_{Li}^V dl_i) - \sum_{i=1}^N w_i s_{Ki}^V \left[\frac{P_{Ki} - P_K}{P_{Ki}} \right] dk_i - \sum_{i=1}^N w_i s_{Li}^V \left[\frac{P_{Li} - P_L}{P_{Li}} \right] dl_i \\ &= \sum_{i=1}^N w_i dx_i^V - I_K - I_L. \end{aligned}$$

By substituting this equation and the definition of aggregate output growth (equation (23)) into the definition of aggregate productivity, we find:

$$dp = \sum_{i=1}^N w_i dv_i - \sum_{i=1}^N w_i dx_i^V + I_K + I_L.$$

We now substitute in equation (21) for firm-level value-added growth:

$$dp = \sum_{i=1}^N w_i (\mu_i^V - 1) dx_i^V + \sum_{i=1}^N w_i (\mu_i^V - 1) \left[\frac{s_{Mi}}{1 - s_{Mi}} \right] (dm_i - dy_i) + I_K + I_L + \sum_{i=1}^N w_i dt_i$$

One of our goals is to ask when aggregate productivity can be expressed as a simple function of aggregate inputs. To do so, we find it useful to rearrange the first term by pulling out the ‘‘average’’ markup and substituting in from equation (24) for primary input growth:

$$\begin{aligned}\sum_{i=1}^N w_i (\mu_i^V - 1) dx_i^V &= (\bar{\mu}^V - 1) \sum_{i=1}^N w_i dx_i^V + \sum_{i=1}^N w_i (\mu_i^V - \bar{\mu}^V) dx_i^V \\ &= (\bar{\mu}^V - 1) dx^V + \sum_{i=1}^N w_i (\mu_i^V - \bar{\mu}^V) dx_i^V + (\bar{\mu}^V - 1)(I_K + I_L).\end{aligned}$$

Using this rearrangement, equation (26) from the text follows immediately:

$$dp = (\bar{\mu}^V - 1) dx^V + R_\mu + R_M + \bar{\mu}^V R_K + \bar{\mu}^V R_L + dt.$$

Appendix B: Using Industry-Level Data

By hypothesis, equation (15) holds at the level of firms, i , in an industry j . For notational convenience, define gross-output-augmenting technology change for firm i by dz_i . Equation (A6) shows that $dz_i = (1 - \mu_i s_{Mi}) dt_i$. We want a measure of (gross-output-augmenting) technology change for the industry: $dz_j = \sum_i w_i dz_i$. However, we estimate equation (15) using industry-level data, so that:

$$(B1) \quad dy_j = \bar{\mu}_j dx_j + R_j + dz_j,$$

where R_j is the sum of the (gross-output) reallocation terms we for industry j : the analogue of equation (26) for gross output. Suppose that R_j is related in (relatively) structural fashion to inputs in industry j and to other influences:

$$(B2) \quad R_j = \alpha_j dx_j + \eta_j,$$

where η_j is uncorrelated (by construction) with dx_j , but may be correlated with aggregate variables such as dx^V . Then an instrumental variables estimation of (B2) would yield (in plim) $\bar{\mu}_j = \mu_j + \alpha_j$, and taking the constant and error terms from that regression we would construct $d\bar{z}_j = dz_j + \eta_j$. Thus one likely explanation of our finding values of μ smaller than one in various industries is that α_j is negative in those industries.³² Note, however, that by itself this issue does not present a problem for estimating technology change. The only divergence between estimated technology change and its true value comes from the presence of η_j .

³² This hypothesis is consistent with a reasonable story. Suppose that industries consist of a number of firms of varying efficiency. At times of low output, the most efficient firms produce almost all of the output. When demand increases, inputs increase systematically at the less-efficient firms. This story implies that *intra*-industry reallocations (as distinct from *inter*-industry reallocations) should be negatively correlated with input use.

Of course, we are interested in the properties of our estimated series of aggregate technology shocks, not technology change in a single industry. However, we discuss the single-industry case to keep the argument clear, since it extends to the aggregate case in fairly straightforward fashion.

We are most interested in the correlation of $d\hat{z}_j$ with various measures of aggregate input and output growth. For concreteness, suppose we are interested in the correlation between $d\hat{t}_j$ and dx^V .

How does the presence of η_j affect this correlation? We find:

$$(B3) \quad \text{corr}(d\hat{z}_j, dx^V) = \frac{\text{cov}(dz_i, dx^V) + \text{cov}(\eta_i, dx^V)}{\sigma_{dx^V} (\sigma_{dz}^2 + \sigma_{\eta}^2 + 2 \text{cov}(dz_i, \eta_i))^{1/2}}.$$

The size of the numerator is affected by the sign of the covariance between η and dx^V . One generally expects this covariance to be positive; indeed we find that result using our industry data.

Of course the size of the correlation also depends on the estimated standard deviation of technology shocks. If η is (weakly) positively correlated with dz_j , then the estimated standard deviation is biased up, and the estimated correlation is biased down. Since the covariance would have to be strongly negative to offset the extra noise coming from the standard deviation, it seems likely that we will estimate a standard deviation of $d\hat{t}_j$ that is too large.

Thus, the upshot is that the correlations we stress are ambiguously affected by our use of industry-level data rather than firm-level data, but our procedure does not introduce any clear bias. It would certainly be preferable to use firm-level data for every firm in the economy over our sample period, but these data do not exist. However, even setting aside our interest in aggregates, it is not at all clear that we would be better off doing this exercise with a small firm-level data set. The reason is that firm-level data are generally contaminated with a great deal of measurement error. If the measurement error is classical, it will wash out in industry-level data or in a large firm-level data set. But in a small firm-level data set, classical measurement error will bias up the estimated standard deviation of technology shocks,

and bias down the correlation (since classical measurement error is equivalent to the case where η is uncorrelated with both dx^v and dt_j).

Note that there is a clear case for instrumenting the right-hand-side variable, input growth, even when one uses industry rather than firm data. IV estimation at least produces a consistent estimate of $\mu_j + \alpha_j$, minimizing the measurement error in \hat{dt}_j . OLS estimation would generally introduce another error by producing inconsistent estimates of $\mu_j + \alpha_j$; the issue, of course, is the famous “transmission problem” of technology change inducing variations in input use. (As argued by Basu and Fernald (1997), the instrumented regressions do not estimate μ alone, because reallocation effects may be correlated with the instruments. In this case, we *want* to capture that correlation.)

We have discussed errors in estimating gross-output technology change, but the quantity we wish to examine is value-added technology change:

$$dt_j = \sum_i w_i^V \frac{dz_i}{1 - \mu_i s_{Mi}}$$

Using industry data, we approximate dt_j by $\hat{dt}_j = \frac{\hat{dz}_j}{1 - \mu_j \bar{s}_M}$.

Even if we correctly estimate $\hat{\mu}_j = \bar{\mu}_j$, using the average markup $\bar{\mu}$ and the average materials share \bar{s}_{Mi} implies (by Jensen’s inequality) that the variance of \hat{dt}_j is less than the true variance of dt_j^V . In some industries, the estimated variance is further reduced by the fact that our estimated $\hat{\mu}_j$ is probably smaller than $\bar{\mu}_j$, since α_j sometimes appears to be negative. On the other hand, as we noted above, the variance of \hat{dt}_j is likely to be biased up because of noise coming from η_j . Thus, there is no clear bias to the estimated variance of dt_j^V .

However, these issues that complicate our estimation of the variance of dt_j^V should not matter for the correlations that are the major focus of the results. With intra-industry reallocations that take the

form of equation (B2), the estimated value-added series typically has the same correlation properties as the estimated gross-output series that we discuss, under the reasonable condition that the distributions of firm-level markups and materials shares are independent of the distributions generating firm-level technology change. (For example, this condition rules out the case where technology change is systematically more cyclical in firms with above-average markups.)

Of course, we examine technology change for the private economy as a whole, not in any one industry j . A similar independence condition says that we do not expect our estimates of the correlations between aggregate technology change and other variables to be biased by using estimated industry-level markups instead of firm-level markups to construct value-added technology change.

REFERENCES

- Barro, Robert J. and King, Robert G. (1984). "Time-separable Preferences and Intertemporal Substitution Models of Business Cycles." *Quarterly Journal of Economics* 99 (November): 817-839.
- Basu, Susanto (1995). "Intermediate Goods and Business Cycles: Implications for Productivity and Welfare." *American Economic Review* 85 (June) 512-531.
- _____ (1996). "Cyclical Productivity: Increasing Returns or Cyclical Utilization?" *Quarterly Journal of Economics* 111(August) 719-751.
- Basu, Susanto and Fernald, John G. (1995). "Are Apparent Productive Spillovers a Figment of Specification Error?" *Journal of Monetary Economics* 36 (December) 165-188.
- _____ and _____ (1997). "Returns to Scale in U.S. Manufacturing: Estimates and Implications." *Journal of Political Economy* 105 (April) 249-283.
- Basu, Susanto; Fernald, John G. and Horvath, Michael T. K. (1996). "Aggregate Production Function Failures." Manuscript.
- Bai, Chong-En; Li, David D. and Wang, Yijian (1997). "The Welfare Significance of Productivity Growth in Chinese State Enterprises." *Journal of Comparative Economics*, forthcoming.
- Baily, Martin N.; Hulten, Charles and Campbell, David. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity (Microeconomics)*, 1992, (1), pp. 187-267.
- Beaudry, Paul and Devereux, Michael (1994). "Monopolistic Competition, Price Setting, and the Effects of Real and Nominal Shocks." Manuscript, Boston University.
- Berndt, Ernst R. and Fuss, Melvin A. (1986). "Productivity Measurement with Adjustments for Variations in Capacity Utilization and Other Forms of Temporary Equilibrium." *Journal of Econometrics* 33 (October/November): 7-29
- Bertin, Amy L.; Bresnahan, Timothy F. and Raff, Daniel M. G. (1995). "Localized Competition and the Aggregation of Plant-Level Increasing Returns: Blast Furnaces 1929-1935." Working Paper 95-11, Reginal H. Jones Center, Wharton School, University of Pennsylvania.
- Bruno, Michael (1984). "Raw Materials, Profits, and the Productivity Slowdown." *Quarterly Journal of Economics* 99 (February): 1- 29.
- 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; Eichenbaum, Martin and Rebelo, Sergio (1995). "Capital Utilization and Returns to Scale." In Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual*.

- Caballero, Ricardo J. and Lyons, Richard K. (1992). "External Effects in U.S. Procyclical Productivity." *Journal of Monetary Economics* 29, 209-226.
- Cogley, Timothy and Nason, James M. (1995). "Output Dynamics in Real-Business-Cycle Models." *American Economic Review* 85 (June) 492-511.
- Cooley, Thomas F. and Prescott, Edward C. (1995). "Economic Growth and Business Cycles." In Thomas F. Cooley, ed., *Frontiers of Business Cycle Research*. Princeton: Princeton University Press.
- Denison, Edward F. (1979). *Accounting for Slower Growth: The United States in the 1970s*. Washington: The Brookings Institution.
- Diamond, Peter A. and Mirrlees, James A. (1971). "Optimal Taxation and Public Production II: Tax Rules." *American Economic Review* 61 (June): 261-78.
- Domar, Evsey D. "On the Measurement of Technical Change." *Economic Journal* 71 (December): 710-729.
- Farmer, Robert and Guo, Jang-Ting (1994). "Real Business Cycles and the Animal Spirits Hypothesis." *Journal of Economic Theory* 63, 42-72.
- Gilchrist, Simon and Williams, John (1996). "Putty-Clay and Investment: A Business-Cycle Analysis." Manuscript, Boston University.
- Griliches, Zvi (1987). "Productivity: Measurement Issues." In John Eatwell, Murray Milgate and Peter Newman (eds.) *The New Palgrave Dictionary of Economics* (New York: Stockton Press).
- Hall, Robert E. (1988). "The Relation Between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96 (Oct.) 921-947.
- _____ (1990). "Invariance Properties of Solow's Productivity Residual." In Peter Diamond (ed.) *Growth, Productivity, Employment* (Cambridge: MIT Press).
- Hall, Robert E. and Jorgenson, Dale W. (1967). "Tax Policy and Investment Behavior." *American Economic Review* 57 (June) 391-414.
- Horvath, Michael T. K. (1995). "Cyclical and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks." Manuscript, Stanford University.
- Hulten, Charles (1978). "Growth Accounting with Intermediate Inputs." *Review of Economic Studies* 45, 511-518.
- 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.

- Judd, Kenneth L. (1995). "The Optimal Tax Rate on Capital Income is Negative." Manuscript, Hoover Institution.
- Katz, Lawrence F. and Summers, Lawrence H. (1989) "Industry Rents: Evidence and Implications." *Brookings Papers on Economic Activity (Microeconomics)*, (1), pp. 209-290.
- Kimball, Miles S. (1995). "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit, and Banking* 27 (November): 1241-77.
- Phelan, Christopher and Trejos, Alberto (1996). "On the Aggregate Effects of Sectoral Reallocation." Manuscript, Northwestern University.
- Rotemberg, Julio J. and Saloner, Garth (1986). "A Supergame-Theoretic Model of Price Wars During Booms." *American Economic Review* 76 (June)390-407.
- Rotemberg, Julio J. and Woodford, Michael (1992). "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy* 100 (December) 1153-1207.
- _____ and _____ (1995). "Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets." In Thomas F. Cooley, ed., *Frontiers of Business Cycle Research*. Princeton: Princeton University Press.
- _____ and _____ (1996). "Imperfect Competition and the Effects of Energy Price Increases on Economic Activity." *Journal of Money, Credit and Banking* (November).
- Shapiro, Matthew D. (1996). "Macroeconomic Implications of Variation in the Workweek of Capital." *Brookings Papers on Economic Activity* (2): 79-119.
- Ramey, Valerie A. and Shapiro, Matthew D. (1997). "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie-Rochester Conference Series on Public Policy*, forthcoming.
- Solow, Robert M. (1957). "Technological Change and the Aggregate Production Function." *Review of Economics and Statistics* 39: 312-320.
- Weitzman, Martin L. "On the Welfare Significance of National Product in a Dynamic Economy." *Quarterly Journal of Economics* 90 (February): 156-62.

Figure 1

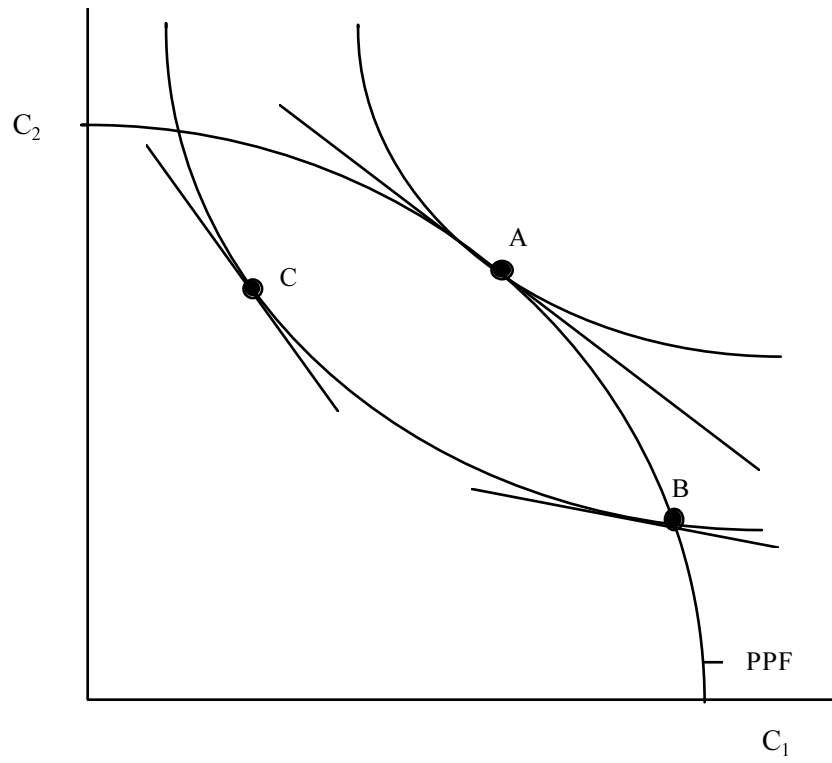
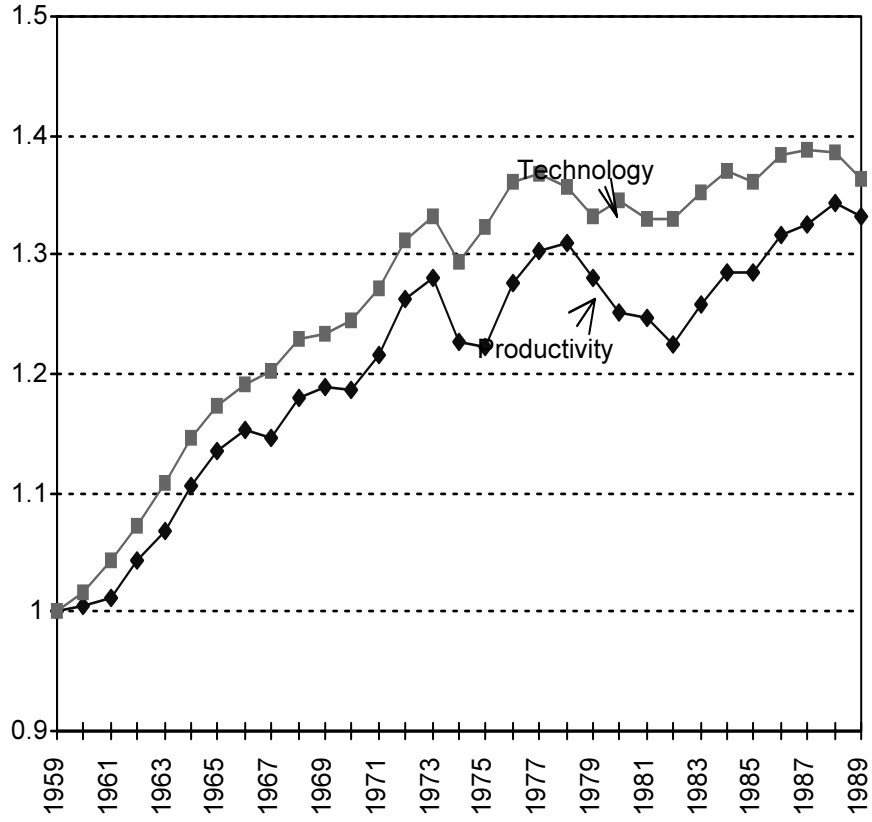


Figure 2
Levels of Productivity and Technology



Note: Levels of the series are calculated by normalizing the levels to 1.0 in 1959 and cumulating estimated growth rates. Growth rate calculations are described in text.

Table 1
Descriptive Statistics

1959-89

	Output Growth dv	Primary Input Growth dx^V	Aggregate Productivity	Corrected Technology Residual (OLS Estimates)	Corrected Technology Residual (IV Estimates)
Mean	3.60	2.35	1.25	1.21	1.11
Standard Deviation	2.84	1.52	2.08	1.62	1.65
Minimum	-2.32	-1.14	-4.05	-2.77	-3.02
Maximum	7.53	5.10	4.28	3.57	3.33

Note: Entries are percentage points per year. Aggregate productivity is the revenue-weighted residual for the private business economy. Instruments are the growth rates in the world price of oil (deflated by the GDP deflator) and government defense spending, and the political party of the president, along with one lag of each.

Table 2

Correlations of Series in Table 1

	Output Growth <i>dv</i>	Input Growth <i>dx^V</i>	Standard Productivity	Corrected Technology Residual (OLS Estimates)	Corrected Technology Residual (IV Estimates)
Output Growth <i>dv</i>	1				
Input Growth <i>dx^V</i>	0.70	1			
Standard Productivity	0.85	0.23	1		
Corrected Technology Residual (OLS Estimates)	0.67	0.05	0.88	1	
Corrected Technology Residual (IV Estimates)	0.53	-0.13	0.81	0.96	1

Table 3
 Reallocation by Component, 1959-89

A. OLS Estimates

	R	R_{μ}	R_M	$\bar{\mu}^V R_K + \bar{\mu}^V R_L$
Mean	-0.082	-0.042	0.11	-0.15
Standard Deviation	0.99	0.63	0.46	0.23
Minimum	-2.65	-1.78	-0.86	-0.62
Maximum	1.79	0.96	1.05	0.60

B. IV Estimates

	R	R_{μ}	R_M	$\bar{\mu}^V R_K + \bar{\mu}^V R_L$
Mean	-0.16	-0.11	0.12	-0.17
Standard Deviation	1.11	0.72	0.51	0.23
Minimum	-3.28	-2.13	-0.99	-0.67
Maximum	1.37	0.94	0.93	0.50

Note: Entries are percentage points per year.

Table 4

Effects of Lagged Technology Improvements on Reallocation

1961-89

	OLS Technology Measure	IV Technology Measure
$\hat{dt}^V(-1)$	-1.06 (0.35)	-0.81 (0.35)
$\hat{dt}^V(-2)$	0.11 (0.35)	-0.13 (0.34)
R^2	0.26	0.20

Note: Dependent variable is reallocation, as defined in the text. Regressions include a constant.

Table 5
Descriptive Statistics by Subperiod

A: 1959-73

	Aggregate Productivity	Corrected Technology Residual (OLS Estimates)	Corrected Technology Residual (IV Estimates)
Mean	2.25	2.26	2.13
Standard Deviation	1.39	0.95	0.93
Minimum	0.08	0.65	0.31
Maximum	4.28	3.57	3.33

B: 1974-89

	Aggregate Productivity	Corrected Technology Residual (OLS Estimates)	Corrected Technology Residual (IV Estimates)
Mean	0.32	0.22	0.14
Standard Deviation	2.22	1.49	1.60
Minimum	-4.05	-2.77	-3.02
Maximum	4.26	2.90	2.86

See note to Table 1.