MARTIN NEIL BAILY
Brookings Institution
ROBERT J. GORDON
Northwestern University

The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power

ALMOST TWO DECADES have now passed since U.S. productivity growth first showed signs of slowing, more than 15 years since the first paper on that topic appeared in this journal.[1] Overall, the slowdown continues with little relief; in the nonfarm business sector the annual growth rate for both output per hour and multifactor productivity was more than 1.5 percentage points slower during 1973–87 than during 1948–73.[2] If the

Baily's work is part of the program of research being carried out by the Center for Economic Progress and Employment at Brookings. The Center is financed by grants from the Ford Foundation and numerous corporate sponsors. Gordon's research is supported by the National Science Foundation. The authors are grateful to Hiranthi de Silva, Hensley Evans, and Gabriel Sensenbrenner for research assistance; to Paul Pieper for helpful discussions and indispensable aid in locating sources of construction data; to Edward Denison for many perceptive suggestions on an earlier draft; and to members of the Brookings Panel and the Northwestern University–Federal Reserve of Chicago Macroeconomics Workshop for helpful comments.

1. William D. Nordhaus, "The Recent Productivity Slowdown," *BPEA, 3:1972*, pp. 493–536.

2. Multifactor productivity growth is computed (see equation 1 below) as the rate of growth of output minus the weighted average of the rates of growth of inputs (capital and labor when output is value added; energy and materials are included otherwise). The weights are the shares of the factors in total cost.

productivity slowdown continues, it must inevitably reduce the ability of the United States to increase its per capita income and wealth, just as it has already resulted in a near-total cessation in the growth of economywide real hourly compensation since 1973.[3] In this sense the productivity growth slowdown might be described as America's greatest economic problem.

Even as economists remain perplexed about the nature of the slowdown, a new puzzle has presented itself. Productivity has recovered strongly in part of the economy while worsening elsewhere. The revival is in manufacturing productivity, where growth in the past half-decade has been almost enough to wipe out the entire 1973–87 deficit compared with 1948–73. The worsening trend is in nonfarm nonmanufacturing, where output per hour has grown at close to a zero rate on average since 1973, while multifactor productivity growth has been *negative*.

Key Measurement Problems in the Aggregate and Industry Data

One goal of this paper is to answer the perplexing question that arises again and again: "Can measurement errors 'explain' all, some, or none of the post-1973 U.S. productivity growth slowdown?" Our answer is "some, but not much." On the basis of hard evidence and some speculation, we conclude that measurement errors are unlikely to explain more than one-third of the post-1973 slowdown in nonfarm business private output per hour—about 0.5 percentage point of the total slowdown of 1.6 percentage points.

But the paper is about far more than measurement and the productivity slowdown. We have both an educational and investigative purpose. We begin by examining the latest official measures of productivity growth and the slowdown and explaining the steps required to develop productivity measures at the aggregate and industry levels.

We emphasize the distinction between aggregate and industry data because many sources of measurement error do not help explain the

3. The annual growth rate of business sector real compensation per hour during 1973–87 was 0.3 percent, as contrasted to a rate for per capita real disposable income of 1.4 percent (see *Economic Report of the President, February 1988*, tables B-46 and B-27). The serious implications of the productivity problem for the American standard of living have been disguised in part by the movement of women into the labor force, but this rise in labor-force participation has already slowed and in any case cannot be sustained forever.

post-1973 slowdown. To affect the estimated size of the overall slowdown, a given measurement problem must have caused *aggregate* output growth to have been understated more (or aggregate input growth overstated more) after 1973 than before.

In contrast, a measurement problem that understates the output growth of a particular industry, without any direct implications for aggregate output, merely reshuffles measurement of productivity growth among industries. For instance, an understatement of output growth in the railroad freight industry would be a pure industry phenomenon, since all of railroad freight output is an intermediate good. But an understatement of real consumer purchases of air transportation would contaminate measures of productivity growth both in the airline industry and in the economy as a whole. Many debates about productivity measurement concern the validity of industry measures. Our findings imply more for the industry allocation of productivity growth than for the overall magnitude of the slowdown.

A useful way to summarize this point is to establish four quadrants on a simple grid as a classification of actual or possible measurement errors.

Affects aggregate economy, contributes to post-1973 slowdown	Affects aggregate economy, same effect before and after 1973
No aggregate effect, contributes to post-1973 slowdown for an industry	No aggregate effect, same effect before and after 1973 for an industry

Only measurement issues that qualify for the northwest quadrant will help to explain the productivity growth slowdown. Issues relating to the southwest quadrant merely reshuffle the industry allocation of productivity change. Issues entering the quadrants in the eastern half of the table could create a secular bias in productivity at the aggregate (northeast) or industry (southeast) levels, but have no implications for the slowdown.

The tour of the basic data and the measurement primer are followed by the core of the paper, our investigation into "The Case of Multiple Mismeasurement." Almost everywhere we look, we find culprits, some of which imply only a misallocation of productivity growth at the industry level, but others of which contribute a partial explanation of the aggregate slowdown. Both types of mismeasurement are important. Misallocating output by industry makes it difficult to diagnose the causes of the slowdown. Inferences about the importance of capital formation or education may be falsified by errors in industry output or input data.

For a mismeasurement to help explain the aggregate slowdown, it is necessary that something be different about how output or input is being measured post-1973. It is not enough just to point a finger at a perennial problem like "unmeasured quality improvements." Instead, we must show that official measures have gotten less accurate. Such a conclusion does not require that measurement methods at the official agencies, particularly the Bureau of Labor Statistics (BLS) and Bureau of Economic Analysis (BEA), have actually deteriorated. Rather, the economy may have changed in ways that make conventional measurement methods less accurate. Our review of the ways in which mismeasurement has become a greater problem in the slowdown period features four main themes: computers, construction, convenience, and consistency.

Main Themes: Computers, Construction, Convenience, Consistency

The explosion of computer power during the past decade is at the heart of the economy's movement toward activities that are hard for conventional methods to measure and provides a plausible reason why measurement errors might have overstated the extent of the post-1973 slowdown. The construction industry's 40 percent productivity decline strains credulity. Convenience, an unmeasured product increasingly available in the retail sector, could in principle involve additions of value to household well-being as important as those contributed by the computer. Finally, the presence or absence of consistency among alternative sources of data provides a way of identifying measurement problems worthy of further scrutiny.

COMPUTER POWER

An important part of the revival of manufacturing productivity and part of the reason for weak labor productivity and negative multifactor productivity (MFP) growth in nonfarm nonmanufacturing is that official data show enormous productivity gains in the manufacture of computers but apparently little productivity improvement in their use. Stemming from the introduction (beginning in early 1986) of a hedonic price index for computers into the U.S. national income and product accounts (NIPA), all official U.S. output, productivity, and MFP data now incorporate the effects of this computer price index, which declines during 1969–87 at an annual rate of 14 percent. This large imputed price decline yields an annual average increase of productivity in the non-electrical machinery industry of nearly 12 percent a year during 1979–87 and an annual average increase of real computer investment over the same period of 24 percent.[4] Introducing numbers like this into an otherwise sluggish economy does startling things to the data, especially for such ratios as the relative price of capital, investment-to-GNP, and capital-to-GNP.

The phrase "computer power" rather than "computers" better describes this element of the productivity story, because the hedonic deflator for computers amounts to measuring the price of a computer "calculation" rather than a computer "box." A key issue for this paper, therefore, is to explore why official data seem not to be showing the payoff from investments in computer power. What has all that computer power been doing, and where is the "black hole" into which all those computers are disappearing?

We explore the computer issue in three ways in this paper. First, we examine conceptually what activities are performed by computers, how such activities should be treated in principle, and how in practice they might be missed altogether in standard data, be undertaken for reasons that are privately profitable even though socially unproductive, or lead to short-term inefficiency of resource use. Second, we look at how computers have affected economywide data and how current index number methodology may lead to incorrect aggregate capital and output series. Third, we make a detailed case study of a key service sector, finance, insurance, and real estate (FIRE), in which computers have had a big impact.

4. Nonelectrical machinery includes the computer industry. For investment, the figures given are the fixed-weight deflator and real investment for the "Office, Computing, and Accounting Machinery" (OCAM) component of producers' durable equipment, NIPA, tables 7.13 and 5.7, respectively. These refer to all office machinery (other than photocopy), of which computer systems (processors and peripherals) now make up a share of about 80 percent. The annual rate of change in the deflator for computer systems is about −17 percent and for computer processors about −20 percent.

CONSTRUCTION

The computer quagmire is only the newest and most dramatic of the measurement issues intrinsic to the more general productivity puzzle. An old perennial is the deflation of structures and the measurement of construction output. The NIPA measure of construction output per hour fell in absolute terms almost 40 percent between 1967 and 1987; throughout 1982–87 it was lower than it was in 1948. We undertake a second case study to evaluate the extent of measurement errors in construction, putting together a wide variety of data on output, prices, and quality, and exhibiting a highly suggestive contrast between U.S. and Canadian productivity.

CONVENIENCE

In searching for shifts in the economy that might have made productivity more difficult to measure after 1973, we are struck by the pervasive emergence of consumer services that offer improved convenience, from the suitably named 24-hour "convenience" stores to fast food stores creating millions of "McJobs" to extended supermarket hours to automatic-teller machines. Mismeasurement in price indexes for consumer services may be the single most fruitful area in which to search for errors that have implications for the aggregate (as opposed to industry) productivity slowdown, simply because consumer purchases of services now amount to fully 35 percent of GNP, more than all nonconsumption components of GNP combined. Our third case study, of retail trade, shows that since 1972 an enormous gap has emerged between productivity growth in the food and nonfood components, with stagnation in the former and steady growth in the latter at almost the same pace as in manufacturing. We present some intriguing detailed data on the supermarket industry showing that firms may have deliberately taken steps that reduced measured labor productivity in order to produce more convenience, variety, and quality.

CONSISTENCY

Inconsistency in two measures of the same concept waves a flag to mark the potential incidence of measurement problems. For example, the BLS develops some of its own industry productivity series that are independent of the NIPA value-added measures of industry output. And many industry associations produce output or productivity data from their own surveys. These series are sometimes at variance with output and productivity measures based on NIPA data. Sometimes the inconsistencies are resolvable without postulating an error in either series, but we have found in our case studies that investigating the inconsistencies has pointed to major measurement problems. For example, the stark contrast between U.S. and Canadian measures of construction productivity strengthens the case against the official U.S. series. As another example, the sharply different NIPA and BLS measures of output growth in the airline industry led us to the use of the CPI for air transportation as the underlying culprit; it vastly exaggerates inflation in airline fares by neglecting to take account of the introduction of discount fares.

Other Themes of the Study and Its Limitations

Our main measure of productivity is average labor productivity (ALP), so that measurement errors in computing the labor input are as important as those for output. Hours of work are measured subject to some error, but more important is the discrepancy between measured labor hours and an economic concept of labor input into a production function. An hour of work by a teenager with a high school education is not the same labor input as an hour of work by an experienced, mature, college-educated worker. Standard productivity series ignore this difference.

Although we focus most attention on average labor productivity, we also report multifactor productivity series. Our greater attention to ALP simply reflects data availability, since MFP measures require scarce data on capital input by industry of use; however, the measurement of capital goods prices is important regardless of whether ALP or MFP is at issue. Capital goods are part of GNP, the capital stock is used directly in the calculation of MFP, and capital output per hour is an important determinant of ALP. This paper draws on Robert Gordon's argument that inflation has been overstated and real output understated for the capital goods sector, with the result that growth in aggregate real GNP, ALP, and the capital stock have all been understated. A central implication is that the computer price deflator is only the tip of the capital goods iceberg; numerous other capital goods have exhibited dramatic declines in nominal or real prices, including the single largest category of producers' durable equipment, communications equipment, for which no government price index is computed by either the BEA or BLS.

A CAVEAT

Our goal is to focus attention on conceptual issues, particularly regarding computers and convenience, and to demonstrate in our case studies that measurement can indeed be improved. We are interested not just in pointing at culprits but in suggesting where current methods of data collection can be improved and in setting down a research agenda. Through it all, we stress the intrinsic interest and importance of measurement issues themselves. We do not suggest that they will solve the productivity slowdown puzzle in its entirety.

PLAN OF THE PAPER

The paper is divided roughly into thirds. The first third contains our broad-brush introduction to the official data at both the aggregate and industry levels, starting with a report on the latest news about the aggregate productivity slowdown in both ALP and MFP. We review the basic arithmetic of GNP measurement by final demand, income, and industry product originating, to identify the steps in measurement most likely to introduce errors and review productivity performance by industry and its implications. The middle part of the paper concentrates on labor and capital input. After reviewing recent findings by others on changes in labor quality, we turn to our core topic of measuring the output and capital input of producers' durables, with special attention to computers and other products where technical progress has been rapid. The last third of the paper consists of our four case studies. The first, for finance, insurance, and real estate, treats an industry where the influence of the computer has been pervasive; the second covers construction; the third, retail trade, where convenience is an issue; and the fourth, transportation, where we find a lack of consistency in government measures. The paper ends with some ballpark estimates of the overall impact of measurement errors in explaining the aggregate productivity slowdown and with suggestions for ways in which government statistical agencies could improve their methods.

Table 1. Average Annual Aggregate Productivity Growth, 1948–87, Selected Periods
Percent per year

Measure	*1948–73*	*1973–79*	*1979–87*	*1973–87*	*Change, 1948–73 to 1973–87*
Output per hour					
Business	2.94	0.62	1.32	1.02	−1.92
Nonfarm business	2.45	0.48	1.11	0.84	−1.61
Manufacturing	2.82	1.38	3.39	2.52	−0.30
Nonmanufacturing	2.32	0.16	0.33	0.25	−2.07
Multifactor productivity					
Business	2.00	0.10	0.61	0.39	−1.61
Nonfarm business	1.68	−0.08	0.45	0.22	−1.46
Manufacturing	2.03	0.52	2.56	1.68	−0.35
Nonmanufacturing	1.55	−0.29	−0.28	−0.30	−1.85

Sources: Data for 1948, 1973, and 1979 taken from U.S. Department of Commerce, Bureau of Economic Analysis, *The National Income and Product Accounts of the United States, 1929–82, Statistical Tables* (Government Printing Office, September 1986), tables 6.2 and 6.11. Data for 1987 taken from Department of Commerce, Bureau of Economic Analysis, *Survey of Current Business*, vol. 68 (July 1988), tables 6.2 and 6.11. Multifactor productivity taken from Department of Labor, Bureau of Labor Statistics, "Multifactor Productivity Measures, 1987," *News*, September 30, 1988.

Aggregate Productivity and Its Breakdown by Industry

Table 1 exhibits alternative aggregate productivity series, including the BLS's MFP figures. All of the series show a large growth slowdown after 1973 and a partial recovery after 1979.[5] The recovery in manufacturing has taken the growth rate for that sector since 1979 well above its 1948–73 average, while productivity growth in the nonfarm nonmanufacturing sector has been stuck near zero since 1973. The post-1973 slowdown in manufacturing has almost disappeared, even when the years 1973–79 are included, as shown in the right-hand column in table 1, while the nonfarm nonmanufacturing slowdown exceeds 2 percentage points a year. Business productivity growth exceeds that of the nonfarm business sector and, because the shift of resources out of farming is now very small, the implication is that farm productivity has been growing at a relatively fast rate since 1973, especially since 1979.

The relationship between growth in ALP and growth in MFP has been well known since the early studies by Robert Solow and Edward Denison

5. We do not claim that 1979 was necessarily the turning point, but movements of the economy into and out of recessions in the years 1980–83 disqualify any year in that interval from consideration as an alternative candidate without an explicit regression analysis of cyclical productivity effects.

three decades ago. Designating real output by Q, labor input by L, capital input by K, and the income share of labor by α, we have:

$$(1) \quad d\ln(ALP) = d\ln Q - d\ln L, \text{ and}$$
$$d\ln(MFP) = d\ln Q - \alpha d\ln L - (1 - \alpha)d\ln K$$
$$= d\ln Q - d\ln L - (1 - \alpha)(d\ln K - d\ln L).$$

Because the MFP growth series in the bottom section of table 1 subtracts growth in capital input as well as in labor input from growth in output (applying the appropriate weights as in equation 1), and because capital input throughout the postwar period has grown more rapidly than labor input, each figure in the bottom section of table 1 is smaller than the corresponding figure in the top section. The MFP data confirm that since 1979 manufacturing productivity growth has exceeded that before 1973, although a small slowdown remains for the full period 1973–87. But in the nonfarm nonmanufacturing sector, MFP growth is negative, and it is just barely positive over 1973–87 for the nonfarm business sector as a whole.

THE SIMPLE ARITHMETIC OF GDP ESTIMATES

The NIPA in the United States rely on final sales of goods and services as the primary source of information on output. GNP is the sum of consumption and investment expenditures, government purchases, and net exports. Subtracting net foreign income yields GDP, and subtracting the production of government (in the form of payments to employees) gives private GDP. Compensation of employees in nonprofit organizations and the imputed rent paid on owner-occupied housing are included in final sales, so that when these figures are subtracted off, the resulting aggregate becomes GDP originating in the business sector. The specific magnitudes of these deductions from GNP are shown in table 2.[6]

6. From Jerome A. Mark, "Problems Encountered in Measuring Single- and Multifactor Productivity," *Monthly Labor Review*, vol. 109 (December 1986), table 1, p. 5. Original data from Bureau of Economic Analysis, U.S. Department of Commerce. Government enterprises are included in business output for the calculation of labor productivity, but excluded in the multifactor productivity calculations because of a lack of capital data from this sector. The concept of private business output displayed in this table is that used by the BLS and, by excluding owner-occupied housing and the statistical discrepancy, does not correspond to the NIPA concept of business product displayed in NIPA tables 1.12 and 1.13.

Table 2. Categories of Output Subtracted from GNP to Give Private Business Output, 1985 Values

Category		*Billions of 1982 dollars*	*Percent*
GNP		3,585.2	100.0
Less:	General government	355.5	9.9
	Owner-occupied housing	209.4	5.8
	Rest of world	37.0	1.0
	Households, nonprofits	140.0	3.9
	Government enterprises	43.9	1.2
	Statistical discrepancy	−5.0	−0.1
Equals:	Private business output	2,804.4	78.2

Source: Jerome A. Mark, "Problems Encountered in Measuring Single- and Multifactor Productivity," *Monthly Labor Review*, vol. 109 (December 1986), table 1, p. 5.

In computing current-dollar GNP from final sales, some output in the underground and illegal economies will inevitably be missed, and any such omissions will flow through the above subtractions to contaminate current-dollar business GDP. It is likely that small businesses and self-employed persons conceal part of their legitimate activity in order to avoid taxes, and the underground economy is reputed to have grown rapidly in recent years, leading to a downward bias in GDP growth.[7] We do not find the evidence presented on this issue to be terribly persuasive, however, at least with respect to its implications for productivity.[8] First, even if the underground economy has grown, this probably would have raised productivity growth, because labor input is concealed as well as output, and the underground activities are surely low-productivity ones.

7. Edgar L. Feige, "How Big Is the Irregular Economy?" *Challenge* (November–December 1979), pp. 5–13; and Feige, "The Theory and Measurement of the Unobserved Sector of the Economy: Causes, Consequences and Implications," paper delivered at the 93rd annual meeting of the American Economic Association, September 6, 1980; Vito Tanzi, ed., *The Underground Economy in the United States and Abroad* (Lexington, 1982).

8. Edward Denison has critiqued the underground economy literature. → Edward F. Denison, "Is U.S. Growth Understated Because of the Underground Economy? Employment Ratios Suggest Not," *Review of Income and Wealth* (March 1982), pp. 1–16; Denison, "Accounting for Slower Economic Growth: An Update," paper presented to the Conference on International Comparisons of Productivity, American Enterprise Institute, Washington, D.C., September 30, 1982; and Denison, *Trends in American Economic Growth, 1929–82* (Brookings, 1985) (see pp. 56–57 for additional references to the literature). Denison has reminded us that the BEA does adjust GNP based upon an estimate of underreporting from the Internal Revenue Service. In addition, it is correct to exclude most illegal activities, such as drug selling, from GNP.

Second, the sharp reduction in marginal tax rates in the 1980s should have shifted both output and employment into the recorded economy. There is no sign of such a shift in GNP growth or the employment-to-population ratio. We will assume, therefore, that the nominal value of business GDP is known with a relatively high degree of accuracy.

DEFLATION ERRORS AT THE AGGREGATE LEVEL

A greater potential for error is introduced when the nominal value of GDP is deflated to give real business GDP using about 800 different commodity deflators. Most of these deflators are components either of the consumer price index (CPI) or the producer price index (PPI). In some cases, however, there is no genuine deflator and for some commodities, real production is inferred from the number of people employed in producing them. Some services such as banking fall into this category. The breakdown of 1985 private business output by type of deflation method is shown in table 3.[9] Factor payments are used as a deflator for 8.5 percent of output, leaving about 92 percent of the total that has a legitimate price deflator.

Even in cases where there is a legitimate deflator, however, the split between real output change and inflation is not necessarily made correctly. Price indexes for durable goods may miss quality improvements taking the form of improved performance, reduced energy use, and a lower frequency of repair. Price reductions that often occur in the first few years after the introduction of a new product may be missed through its late incorporation into the CPI or PPI. Compared with a sizable body of research on the deflation of durable goods, the study of errors in service prices is just beginning. One reason for slow reported productivity growth in services is that increases in computers and support staff may be providing an improved quality of services, or entirely new services, that the service deflators are not capturing. We also believe that there may be a widespread failure of existing deflators to capture the upgrading of service quality that occurs when, for instance, supermarkets offer a broader selection, a barber installs air conditioning, or a hotel equips every bathroom with a phone and TV set. As the four-quadrant diagram in the introduction makes clear, however, if the poor

9. From Mark, "Problems Encountered in Measuring Single- and Multifactor Productivity."

Table 3. Real Private Business Output, by Type of Deflation Method, 1985 Values

Item	*Billions of 1982 dollars*	*Percent*
Portion deflated by compensation or cost indexes	238.5	8.5
Nonresidential structures	152.2	5.4
Other	86.3	3.1
Portion deflated by price indexes	2,565.9	91.5
Total	2,804.4	100.0

Source: Mark, "Problems Encountered in Measuring Single- and Multifactor Productivity."

deflation of durable goods output or of services output contributes to the explanation of the ALP growth slowdown, the shares of these commodities must have risen as a share of total business product, or unmeasured quality improvement must have increased in importance. In the case of services, these possibilities seem plausible, both because the overall share of services has risen and because some of the poorly measured areas, such as financial services, have seen much innovation in recent years.

FROM AGGREGATE INCOME AND PRODUCT TO INDUSTRY PRODUCT ORIGINATING

Because of the identity between income and product, GNP is also equal to the sum of all income payments. Data on income are available from tax and Securities and Exchange Commission records, so after allowing for capital consumption allowances, GNP can also be computed from income. The reported statistical discrepancy in the two calculations usually lies in the range ± 0.2 percent of GNP. That discrepancy, however, is what remains after BEA has done its best to bring about consistency. Income and product data are adjusted based upon areas where BEA judges there are reporting errors.

Once income payments associated with government, nonprofit organizations, and foreign activities are subtracted from total income, what remains is income generated in the business sector, which should be equal to the GDP originating in that sector. The BEA's attempts to allocate this income by industry encounter difficulties, because many companies span several industries. The allocation of most employee compensation can be made reasonably well by industry, although head office staff do provide services to divisions and plants and sometimes vice versa. Interest income is allocated, often crudely, to the industry designated as the primary industry of a company, and profits and depreciation must be assigned by industry without adequate data. There is thus a potential for error in the allocation of current-dollar business GDP into its industry components.

DEFLATION AT THE INDUSTRY LEVEL

The potential for error is much greater for *real* GDP by industry, however, than for the current-dollar values. Deflating value added requires estimates of both prices and quantities of intermediate goods and services. The NIPA were set up to measure final goods and services production, rather than intermediate production, and the data base reflects this. In practice the survey coverage of prices of intermediate goods is quite limited, and the quantities of purchases and sales of intermediate goods are not known from year to year, so extrapolations are made from census years. For the manufacturing sector, the annual survey does provide more frequent information on product flows, and the Census Bureau develops its own estimates of value added. Unfortunately, the survey does not ask about purchased services, so there is no direct comparison between income and value added even in manufacturing. Moreover, the survey does not ask about the composition of product purchases, except for energy.

Industry Productivity Trends

Average labor productivity for the major industries of the economy is shown in table 4 for various subintervals over 1948–87. Productivity figures for government, nonprofit organizations, and private households are given for completeness; growth rates in these sectors would be zero except for mix effects and quirks in the way the numbers are put together. Table 4, like table 1, dates the slowdown at 1973 and compares performance pre- and post-1973.[10] The post-1973 decline in growth in the

10. This treatment conceals the fact that ALP growth slowed in the business sector after 1965, associated with slower growth in mining and a large decline in construction productivity. The rest of the economy had no slowdown prior to 1973. See table 5 below.

Table 4. Average Annual Growth in GDP per Hour, Major Sectors of the U.S. Economy, 1948–87, Selected Periods

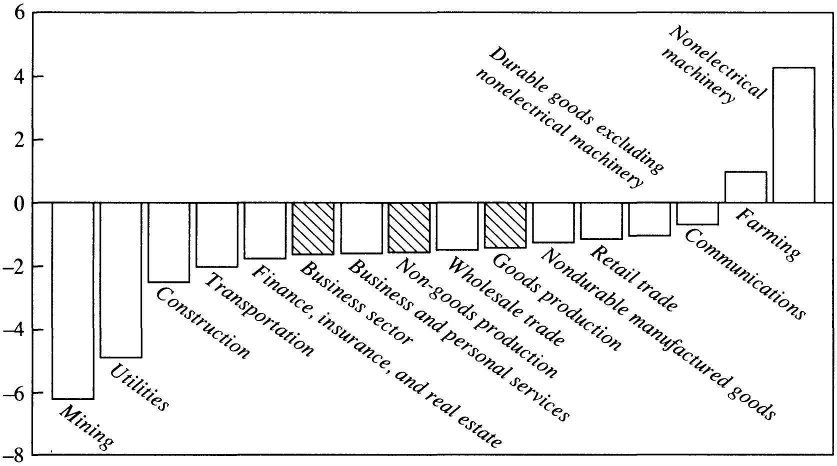
Percent per year

Sector	*1948–73*	*1973–79*	*1979–87*	*1973–87*	*Change, 1948–73 to 1973–87*
Business	2.88	0.63	1.36	1.05	−1.83
Goods-producing industries	3.21	0.55	2.39	1.60	−1.61
Farming	4.64	3.09	6.86	5.22	0.58
Mining	4.02	−7.05	2.34	−1.79	−5.81
Construction	0.58	−1.99	−1.67	−1.80	−2.38
Manufacturing	2.87	1.43	3.49	2.61	−0.26
Durable goods excluding nonelectrical machinery	2.56	1.12	2.09	1.67	−0.89
Nonelectrical machinery[a]	2.03	0.70	11.54	6.76	4.73
Nondurable goods	3.40	1.90	2.13	2.03	−1.37
Non-goods-producing industries	2.49	0.73	0.66	0.69	−1.80
Transportation	2.31	1.06	−0.50	0.17	−2.14
Communications	5.22	4.25	5.09	4.73	−0.49
Electricity, gas, and sanitary services	5.87	0.05	1.44	0.84	−5.03
Trade	2.74	0.76	1.68	1.28	−1.46
Wholesale	3.14	0.10	2.39	1.40	−1.74
Retail	2.40	0.87	1.21	1.06	−1.34
Finance, insurance, and real estate	1.44	0.28	−1.15	−0.54	−1.98
Business and personal services	2.17	0.34	0.36	0.35	−1.82
Government enterprises	−0.15	0.94	−0.15	0.32	0.47
General government[a]	0.21	−0.28	0.37	0.09	−0.12
Nonprofit organizations[b]	0.31	−0.88	−0.32	−0.56	−0.87
Employment in private households[c]	−0.35	−0.63	1.98	0.85	1.20
Rental value of owner-occupied housing	n.a.	n.a.	n.a.	n.a.	n.a.

Sources: Hours and GDP from Bureau of Labor Statistics data except as noted.
n.a. Not available.
a. GDP and hours for 1948, 1973, and 1979 from NIPA; for 1987 from *Survey of Current Business*, vol. 68 (July 1988).
b. GDP from NIPA and *Survey of Current Business*. Hours from *Survey of Current Business* and BLS estimates of military hours.
c. GDP from NIPA and *Survey of Current Business*. Hours from BLS.

business sector was widespread, a point illustrated more vividly by figure 1. All the industries except agriculture and nonelectrical machinery experienced slower growth post-1973 than pre-1973. The ubiquitous nature of the slowdown strongly suggests that the slowdown is real and not just the product of measurement errors. Sectors where productivity is relatively well measured, such as mining, utilities, and nondurable

Figure 1. Change in Labor Productivity Growth, 1948–73 to 1973–87, by Industry
Percent change



manufacturing, all experienced slowdowns. The slowdown in goods-producing and non-goods-producing industries was about the same. The industries with questionable real output data, such as construction, transportation, FIRE, and services, have somewhat larger slowdowns than average, but smaller than mining and utilities. Thus in the next section when we eliminate the poorly measured industries and keep only the well-measured ones, the slowdown is still clearly visible.

At first glance, the industry productivity data suggest that the increase in energy prices in 1973 had an effect on productivity. Mining and utilities, two of the industries most heavily affected by the energy crisis, had the biggest post-1973 slowdowns. Transportation, too, had a major slowdown. On closer inspection, however, the impact of energy is not so clear. Both mining and utilities had begun to slow down before 1973. The depletion of easily available oil reserves in oil extraction, health and safety regulation in coal mining, and the depletion of innovation and returns-to-scale opportunities in electric power, together with the effect of environmental regulations, were reducing growth before the energy crisis hit.[11] In the transportation sector, too, the timing seems a little off.

11. On oil reserves, see William D. Nordhaus, *The Efficient Use of Energy Resources* (Yale University Press, 1979); and Nordhaus, "Oil and Economic Performance in Industrial Countries," *BPEA, 2:1980,* pp. 341–400. On electric power, see Robert J. Gordon, "The

This sector slowed less after 1973 than it did after 1979, a period when energy prices began to come down.[12] So on closer examination, the fact that energy-intensive sectors had post-1973 slowdowns provides less compelling evidence than seemed at first.

Table 4 indicates that growth did make a partial recovery during 1979–87. Growth was substantially faster after 1979 than during 1973–79 in many of the major sectors of the economy, particularly in the goods-producing industries. And because the two industries where the slowdown intensified after 1979 were transportation and FIRE, both of which have measurement problems, it is possible that the partial recovery of measured growth after 1979 might be understated in the official data because of unmeasured output gains.

THE EFFECT OF REMOVING THE PROBLEM CHILDREN

One way to show how much effect there has been from industries with measurement problems is to remove their value added and hours from the total—exactly what is done already when government, nonprofit organizations, and owner-occupied housing are removed. Table 5 shows the effect of removing the problem or controversial industries, either singly or in combination.[13] Regardless of which industries are removed, the existence of a post-1973 slowdown remains clear, and the 1973–79 period was one of strikingly weak growth. In other ways, however, removing these industries makes a big difference.

One such difference is revealed by separating out the period 1965–73, for which the removal of the construction industry has an important

Productivity Slowdown in the Steam-Electric Generating Industry" (Northwestern University, February 1983); Paul L. Joskow and Nancy L. Rose, "The Effects of Technological Change, Experience, and Environmental Regulation on the Construction Cost of Coal-burning Generating Units," *Rand Journal of Economics,* vol. 16, no. 1 (Spring 1985), pp. 1–27; and Martin N. Baily and Alok K. Chakrabarti, *Innovation and the Productivity Crisis* (Brookings, 1988), pp. 67–85.

12. The statement in the text is based on the NIPA data compiled by the BEA. As we show in our case study below, much of the measured slowdown in NIPA transportation productivity is fictitious. While trucking is still a problem, reflecting the 55-mile speed limit and end of the one-time effect of building of the interstate highway system, productivity growth in railroads and airlines has been buoyant. We show that ALP in both railroads and airlines more than doubled during 1967–86, even with no allowance for the effects of computers on the quality of output.

13. The first number in the first column of table 5 is the famous figure of 3.2 percent, the basis of the Kennedy-Johnson anti-inflation "guideposts." After two decades of data revisions, this number is still inviolate.

Table 5. Real Business Sector Labor Productivity Growth Omitting Selected Industries, 1947–87, Selected Periods

Percent per year

Item	*1948–65*	*1965–73*	*1973–79*	*1979–87*	*1948–73*	*1973–87*	*Change, 1948–73 to 1973–87*
Business total[a]	3.23	2.06	0.64	1.30	2.85	1.02	−1.83
Less: Services	3.35	2.13	0.78	1.65	2.96	1.28	−1.68
FIRE	3.27	2.08	0.60	1.50	2.89	1.12	−1.77
Construction	3.19	2.70	0.86	1.51	3.03	1.23	−1.80
Transportation	3.34	2.03	0.62	1.40	2.92	1.07	−1.85
Nonelectrical machinery	3.26	2.05	0.64	0.94	2.88	0.81	−2.07
Services and FIRE	3.41	2.15	0.72	1.93	3.00	1.41	−1.59
Services, FIRE, construction, and transportation	3.54	2.99	0.99	2.51	3.37	1.86	−1.51
All of above	3.61	3.02	1.00	1.97	3.42	1.55	−1.87

Sources: Nonelectrical machinery for 1948, 1965, 1973, and 1979 taken from NIPA; for 1987, from *Survey of Current Business,* vol. 68 (July 1988). All other figures taken from BLS data.

a. Built up from industry data by authors.

effect on the pattern of productivity for the business sector as a whole. Over this interval, output per hour in the total business sector rose 1.17 percent a year more slowly than in the pre-1965 period. Removing construction reduces this relative slowdown more than half, to only 0.49 percent a year. By contrast the white-collar industries have their most important impact after 1979. Removing the white-collar sectors of FIRE and services lifts output per hour to almost 2 percent in the remaining sectors of the business economy. If the other problem industries of construction and transportation are removed, the slow growth in the remaining industries looks like only a temporary problem, with a growth rate of 2.51 percent during 1979–87—a rate that is within shouting distance of the 1948–73 average of 3.37 percent.

The calculations reported in table 5 should be evaluated cautiously, because the industries we are omitting, particularly services, FIRE, and transportation, produce part of their output for intermediate use. Construction and nonelectrical machinery produce primarily for final output. If services, FIRE, and transportation have had real productivity problems, then the table indicates the impact of these on the aggregate. If they have had only measurement problems, then the impact of these on aggregate productivity is less than is indicated in table 5.

IS MANUFACTURING GROWTH BEING OVERSTATED?

The potential errors in the allocation of total GDP by industry to which we have pointed have led Edward Denison to question the validity of productivity analysis by industry.[14] He argues that the allocation of output by industry is so fraught with error that industry productivity measures are unreliable and should be supplemented by estimates that allocate inputs by sectors of final demand. Without prejudging his input allocation proposal, we believe that he exaggerates the problems with industry productivity measures, which, despite their failings, provide essential tools for analysis and can suggest areas where there is mismeasurement with aggregate implications. Denison also argues that there has been a specific bias in recent years, namely, that manufacturing output and productivity have been overstated and nonfarm nonmanufacturing understated. He gives two main reasons for his view. The first is that there has been a normal historical relation between productivity growth in manufacturing and nonmanufacturing, so that the sudden opening up of a large gap in their growth rates seems suspicious. The second reason involves the effects of the new computer price index, which we discuss in conjunction with capital input measurement issues later on.

Denison is not alone in suggesting an overstatement of the growth in manufacturing output. A recent study by Lawrence Mishel has claimed that both current- and constant-dollar shares of the manufacturing sector have been misestimated.[15] Mishel's first argument is that BEA cooked the books to make the manufacturing share of output constant by introducing a "fudge factor" that lowered manufacturing output about $55 billion (1982 dollars) in 1973 and by lesser amounts in other years. The fudge factor was almost zero by 1979, so it had the effect of raising the rate of growth of output during 1973–79 and helping keep constant the manufacturing share over the period. Second, Mishel points out that from 1979 to 1985 the BEA data indicate that the ratio of manufacturing gross output to purchased inputs rose dramatically, marking a sharp shift

14. Edward F. Denison, *Estimates of Productivity Change by Industry: An Evaluation and an Alternative* (Brookings, forthcoming).

15. Lawrence R. Mishel, *Manufacturing Numbers: How Inaccurate Statistics Conceal U.S. Industrial Decline* (Washington, D.C.: Economic Policy Institute, April 1988).

in the earlier trend. Mishel argues that this change is implausible, particularly because the 1979–85 period was one of a rising dollar and widespread reports of increases in outsourcing by manufacturing companies. He points out that BEA has used only domestic price indexes to deflate purchased intermediate goods, ignoring the likely declines in the prices of imported components. He concludes that the growth of manufacturing value added has been overstated for 1979–85, although he does not have a figure for the alleged error over that interval. If manufacturing productivity were measured in accord with Mishel's argument, it would change the picture considerably. Removing BEA's adjustment factor would lower the 1973–79 productivity growth rate in manufacturing almost to zero. Mishel's 1979–85 argument would result in a reduction in the post-1979 recovery of manufacturing productivity.

BEA has responded to the criticisms by Denison and Mishel.[16] They argue that ad hoc adjustments are essential given the weaknesses of the data on income by industry and the need to match total income with total expenditure. When the NIPA were rebased to 1982, a large discrepancy was found between real GDP by commodity and real GDP calculated from the total of income originating by industry in 1972 and 1973. Since the commodity data are the more reliable, BEA scaled back the industry data, leading to the downward adjustment of manufacturing output, by about 8.7 percent in 1972, 8.8 percent in 1973, and by smaller amounts in other years. Looking at the productivity data supports the need for adjustment. Without the 1973 adjustment, productivity growth in manufacturing would have been 3.77 percent a year during 1965–73 and 0.11 percent a year during 1973–79. This is possible, but unlikely.[17] When it comes to the post-1979 recovery, Mishel's argument is much more solid.

16. Department of Commerce, Bureau of Economic Analysis, "Gross Product by Industry: Comments on Recent Criticisms," *Survey of Current Business*, vol. 68 (July 1988), pp. 132–33. The comments were prepared by Frank deLeeuw and Robert P. Parker.

17. Michael Darby has suggested that an important explanation of the 1973–79 slowdown was that 1971–73 real output growth was being overstated and 1973–79 output growth understated because of price controls. Darby argues that the price controls encouraged companies to downgrade product quality, or at least the quality associated with a particular product category, leading to an upward bias in real output during 1971–73 and a downward bias during 1974–75 as controls were removed and companies restored the old levels of quality. See Michael R. Darby, "The U.S. Productivity Slowdown: A Case of Statistical Myopia," *American Economic Review*, vol. 74 (June 1984), pp. 301–22. Both of the present authors have been skeptical of Darby's argument, for different reasons. For Baily's view, see "A Comment on Michael Darby's Explanation of U.S. Productivity Growth" (Brookings, June 29, 1984). Gordon's study of durable goods prices

BEA concedes that the absence of an import price deflator has introduced bias into manufacturing value added. We predict that future data revisions will show somewhat slower growth of manufacturing productivity during 1979–85.

BEA follows the approach, which we also recommend, that consistency among data sources provides an important check on errors. They note that according to an independent Census Bureau estimate, current-dollar value added in manufacturing grew at 4.7 percent a year during 1972–85, compared with 4.3 percent a year for current-dollar gross product. And the Federal Reserve Board's index of manufacturing industrial production grew 3.0 percent a year over the same period, compared with 2.7 percent for constant-dollar gross product.

BEA does concede that the consistency check is not as close for the pre- and post-1979 subperiods. The alternative data sources indicate a little more growth before 1979 and a little less during 1979–85. These estimates are consistent with the Mishel complaint about import prices, but not his complaint about the 1972 and 1973 adjustments. When the dust settles on this issue it is unlikely that the post-1979 recovery of growth in manufacturing will be eliminated. If the overvalued dollar is the source of the manufacturing revival, then the revival should be reversing itself because the dollar has declined. Preliminary data for 1987 and 1988 indicate that the recovery of growth in productivity in manufacturing is continuing.

Mix Effects

The movement of aggregate productivity is not equal to the average of the movements of the individual industries. The use of aggregate series can provide a misleading view of underlying trends, because there is an aggregation or mix effect. Table 6 shows the shares of the main components of the business sector in output and hours of labor input.

described below collected annual data from the Sears catalog for 68 different products, with multiple models for many products, and controlled for every quality characteristic listed in the Sears printed specification. When these Sears prices were compared product-by-product with the detailed PPIs for the same goods, there was no change in the Sears-PPI ratio evident in the period 1971–73, indicating that any quality deterioration must have been heavily disguised or, more probably, nonexistent. We find it likely that microeconomic adjustment costs impede rapid changes in product quality either up or down in response to temporary pricing distortions.

Table 6. Real Output and Labor Shares and Relative Productivities, 1948–87, Various Years

Percent except as noted

Industry	*1948*	*1973*	*1979*	*1987*
		Share of total output		
Farm	5.1	2.3	2.2	2.4
Mining	8.2	6.3	5.3	3.9
Construction	10.2	8.0	7.0	5.8
Manufacturing	27.1	29.2	28.2	27.6
Transportation	8.7	5.5	5.6	4.5
Communications	1.0	2.4	2.9	3.5
Utilities	1.5	3.5	3.4	3.5
Trade	18.4	20.5	20.7	21.7
FIRE	8.8	10.2	11.1	10.9
Services	10.6	12.1	13.7	16.4
		Share of total hours		
Farm	18.4	5.5	4.5	3.3
Mining	1.9	1.1	1.5	1.0
Construction	5.1	7.1	7.3	7.6
Manufacturing	29.1	32.0	29.6	24.7
Transportation	6.3	4.6	4.5	4.2
Communications	1.4	1.8	1.8	1.6
Utilities	1.0	1.2	1.2	1.2
Trade	22.4	26.0	26.0	26.5
FIRE	3.7	6.1	6.8	8.1
Services	10.8	14.7	16.9	21.8
		Relative labor productivity[a]		
Farm	0.28	0.42	0.49	0.75
Mining	4.42	5.80	3.61	3.89
Construction	2.00	1.13	0.97	0.76
Manufacturing	0.93	0.91	0.95	1.12
Transportation	1.38	1.20	1.23	1.06
Communications	0.75	1.32	1.63	2.18
Utilities	1.45	2.97	2.87	2.89
Trade	0.82	0.79	0.80	0.82
FIRE	2.37	1.66	1.63	1.33
Services	0.99	0.83	0.81	0.75

Source: BLS data.

a. Share of total output divided by share of total hours.

Farming, mining, and construction have declined in importance as proportions of goods output, although construction has maintained its share of hours. Most of the non-goods-producing industries have increased their shares of business sector output, with transportation as the main exception.

To estimate the importance of mix effects during the postwar years, we use William Nordhaus's method of decomposing aggregate growth into the weighted average of the rates of growth in the individual industries plus a mix effect (see table 7). The results for the early years are familiar. The shift of workers off the farm, important early on, had ended by 1973. The results of the post-1973 mix effect calculations are new, however. During 1973–79 the mix effect from the nonfarm sector actually boosted aggregate productivity growth, but during 1979–87 mix effects reduced growth. These results strengthen the idea that productivity growth has made a partial recovery. The fixed-weight average productivity growth rate increased 1.2 percentage points after 1979, compared with only a 0.7 point speed-up in the official data.

The findings are generated largely by the fact that ALP in mining is several times as large as the average for all industries. The increase in energy prices in 1973 brought workers into the industry, and the decline in prices in the 1980s pushed them out again. The mining sector alone accounts for 0.23 out of a positive mix effect of 0.26 during 1973–79. This sector accounts for −0.14 out of a negative mix effect of −0.25 during 1979–86. A secondary effect is that the growth of employment in services has had a negative impact on growth: −0.06 during 1973–79 and −0.13 during 1979–87. However, the growth of employment in finance, insurance, and real estate had a positive mix effect on overall growth, adding 0.08 in both of the two post-1973 periods.[18]

Recapitulation and Preview

This completes the first third of the paper, our broad-brush introduction to the aggregate and industry-level productivity measures, as

18. In earlier work Baily argued that decomposing aggregate multifactor productivity growth rather than labor productivity growth was more consistent with the model of a market allocation of factors of production. The results of this MFP decomposition calculated through 1986, using MFP by industry from the American Productivity Center, confirm what we have just reported. Fixed-weight MFP growth also increases by about a percentage point after 1979. That itself is a striking finding, because the MFP calculations in table 1 made by the BLS indicate that the 1979–86 recovery of MFP growth is only about half as large as the recovery of labor productivity growth. Thus the mix effects are found to be more important when calculated from MFP. Martin N. Baily, "The Productivity Growth Slowdown by Industry," *BPEA, 2:1982,* pp. 423–59.

Table 7. Decomposition of Labor Productivity Growth, Business Sector, 1948–87, Selected Periods

Percent per year

Period	*Fixed-weight average productivity growth*	*Effects of changes in output shares*	*Industry mix effects, farm*	*Industry mix effects, nonfarm*	*Business sector total productivity growth*[a]
1948–53	3.01	−0.00	0.65	−0.09	3.58
1953–65	2.81	−0.04	0.31	−0.07	3.01
1965–73	2.09	−0.27	0.22	−0.00	2.04
1973–79	0.38	−0.09	0.09	0.26	0.64
1979–87	1.53	−0.04	0.06	−0.25	1.29
1948–73	2.62	−0.11	0.35	−0.05	2.81
1973–87	1.04	−0.06	0.07	−0.03	1.01
Change, 1948–73 to 1973–87	−1.58	0.05	−0.28	0.02	−1.80

Source: Hours and output taken from BLS data. For further information regarding decomposition of aggregate growth, see Martin N. Baily, "The Productivity Growth Slowdown by Industry," *BPEA, 2:1982*, pp. 423–59.
a. Equals the sum of fixed-weight average productivity growth and mix effects.

compiled by the official agencies. We turn now to the measurement of input, both labor and capital. After reviewing recent work by others on labor quality, we address conceptual issues in the measurement of producers' durable goods, which matter both as output and as capital input. Here we assess the current debate on the measurement of computer prices in the official data. We shall find that quality change in capital goods involves many products beyond computers that have not been treated adequately in the NIPA, and we examine the implications of improved price indexes for these products.

Measuring Labor Input

The use of labor hours as labor input represents a potential source of mismeasurement. There is tremendous heterogeneity in the labor force, and changes over time in the age, sex, or educational mix of the work force change the average quality of labor hours. In addition, the growth slowdown and the U.S. problems with competitiveness have raised the suspicion that the quality of the U.S. work force has declined. Although up to now most research has indicated that trends in labor-force quality have not significantly affected productivity trends, some, though not all, new research suggests a serious decline in labor quality, so it is worth updating this issue.[19]

Denison finds little difference in the post-1973 shift in trend between the raw total of hours and the adjusted index of labor input. He constructs an index of labor input after adjusting hours worked for the effect of changes in the age-sex composition of the work force, the amount of education, and an adjustment resulting from differences in work-weeks by type of person and occupation. The top panel of table 8 shows the results of Denison's adjustments, which imply that labor input has grown substantially faster than total hours over the postwar period as a result of increases in education, and that the effect of education has been remarkably stable over the full period 1948–82. Changes in the age-sex mix of the population have had a negative impact on growth during 1948–82, with this impact intensifying slightly after 1973. Denison explains about 0.1 percent a year of slowdown with his labor quality adjustment.

Dale Jorgenson, Frank Gollop, and Barbara Fraumeni follow Denison in constructing an index of labor input in which the relative wages of individuals are taken as indicative of relative productivities but differ from Denison in making a much finer breakdown of the work force (81,600 cross-classifications) and in using the Törnqvist index number approach.[20] In their calculations, the total change in labor-force quality reflects not only the partial effects of sex, age, education, employment class, and occupation, but also the interactions among them. This means that the sum of the partial effects is not equal to the total effect. The middle panel of table 8 gives a summary of some of their results, which differ importantly from Denison's. Jorgenson and his colleagues find that all of their five elements of labor quality turned adverse after 1973. Labor quality contributed 0.72 to labor input growth prior to 1973, but

19. Denison, *Trends in American Economic Growth*; Dale Jorgenson, Frank Gollop, and Barbara Fraumeni, *Productivity and U.S. Economic Growth* (Harvard University Press, 1987); Edwin Dean, Kent Kunze, and Larry Rosenblum, "Productivity Change and the Measurement of Heterogeneous Labor Inputs," paper presented at the Conference on New Measurement Procedures for U.S. Agricultural Productivity, March 31–April 1, 1988, Washington, D.C.

20. The breakdown into 81,600 boxes includes a breakdown by industry. The effects of industry shifts are not counted as part of the labor quality adjustment, however. The industry shifts are treated separately in their analysis. Their approach has been criticized, because the underlying wage data are not sufficiently detailed to support a breakdown as fine as the one they use.

Table 8. Alternative Adjustments for Labor Quality, 1948–86, Selected Periods
Percent change per year

Denison: Nonresidential business (potential)

Period	Total weekly hours	Adjustment: Age-sex	Adjustment: Education	Adjustment: Group shifts	Adjusted labor input[a]
1948–73	0.73	−0.24	0.64	0.21	1.34
1973–82	2.02	−0.38	0.69	0.19	2.54
Change	1.29	−0.14	0.05	−0.02	1.20

Jorgenson: Whole economy

Period	Total hours	Adjustment[b]: Sex	Adjustment[b]: Age	Adjustment[b]: Education	Adjustment[b]: Employment classification	Adjustment[b]: Occupation	Adjusted labor input[a]
1948–73	1.01	−0.19	−0.07	0.66	0.17	0.37	1.73
1973–79	1.62	−0.54	−0.34	0.36	−0.22	0.00	1.72
Change	0.61	−0.35	−0.27	−0.30	−0.39	−0.37	−0.01

BLS team: Private business sector

Period	Total hours	Quality index adjustment	Adjusted labor input[a]
1948–73	0.68	0.28	0.96
1973–86	1.44	0.30	1.74
Change	0.76	0.02	0.78

Sources: Edward F. Denison, *Trends in American Economic Growth, 1929–82* (Brookings, 1985), table 3-4; Dale Jorgenson, Frank Gollop, and Barbara Fraumeni, *Productivity and U.S. Economic Growth* (Harvard University Press, 1987), tables 8.1 and 8.6; Edward Dean, Kent Kunze, and Larry Rosenblum, "Productivity Change and the Measurement of Heterogeneous Labor Inputs," paper presented at the Conference on New Measurement Procedures for U.S. Agricultural Productivity, March 31–April 1, 1988, Washington, D.C. Total hours taken from BLS, "Multifactor Productivity Measures 1986."

a. Sum of total or total weekly hours and quality adjustments.

b. Adjustments interact so that their total effect is not simply their sum.

only 0.10 after 1973. According to Jorgenson and his colleagues, labor quality accounts for a slowdown of 0.62 in labor productivity growth.

Edwin Dean, Kent Kunze, and Larry Rosenblum of BLS use an approach different from either Denison's or Jorgenson's. They argue that the relative wages of individuals may reflect factors other than relative productivities. They run regressions to determine the effect of experience (not age) and education on wages and use the estimated coefficients to determine the extent to which a change in the overall levels of experience and education have changed the quality of the labor input. Other variables are controlled for in the regressions but do not contribute to the estimate of labor quality change.[21] The results on labor quality are shown in the bottom part of table 8. The BLS group makes smaller adjustments than Denison does and concludes that there was virtually no reduction in the rate of quality augmentation over time.

What should we make of these differences? The BLS group makes smaller labor quality adjustments than Denison, because it argues that only part of wage differentials translates into productivity effects, but the differences are sufficiently minor to lead both Denison and the BLS group to conclude that labor quality adjustments contribute little to understanding the productivity slowdown. In assessing the Jorgenson results, we stress first that they apply to only half the time period covered by the BLS group. The slowdown in the growth of female labor-force participation in the 1980s and the declining share of teenagers guarantees that an extension of the Jorgenson results to 1986 would yield much smaller age-sex effects. As for the general Jorgenson approach, its advantage is that its rigorous basis in production theory means that its estimate of technical change or the productivity residual can be interpreted cleanly as the shift factor in an aggregate production function. But one can argue that changes in occupation and employment class reflect, at least in part, changes of the economic system rather than changes in intrinsic labor quality. This is part of the old debate about whether productivity adheres to the job or the person; ask any woman who has escaped from low wages and occupational crowding in, say, the textile industry to take a job nearby in durable goods manufacturing.

The most puzzling difference between Denison and Jorgenson involves the effect of education. The most likely explanation is that when Jorgenson and his colleagues use Törnqvist index numbers with current period weights, they pick up the decline in the *return* to education that took place in the 1970s.[22] We are dubious of an approach that interprets a reduction in the return to higher education resulting from a change in the balance of supply and demand as a decline in the quality of existing college-educated workers. After all, we would not want to count an

21. These other variables are geographical region, full-time or part-time, veteran status, and residence in a central city. Different regressions are run by sex, so that wage differences by sex are assumed to reflect productivity differences.

22. Richard B. Freeman, *The Overeducated American* (Academic Press, 1976).

existing Boeing 747 as less capital as a result of a decline in airline profitability that may well be temporary. It would be a mistake, then, to interpret Jorgenson's findings as saying that there was a decline in the rate of accumulation of human capital over the period, when accumulation is measured by increased years or days a year of schooling. Denison and the BLS study show that this was not the case. Further, we know that there has been a sharp revival in the return to education in the 1980s.[23] The Jorgenson approach will doubtless show much less difference between total hours and effective labor input when extended forward in time.

TEST SCORES AND LABOR QUALITY

One possible explanation of the decline in the return to education in the 1970s is that the "quality" of a year of schooling may have declined, perhaps because students were not learning as much. An important observation that fueled concern about the size of the payoff from economywide increases in years of schooling was the decline in scholastic aptitude test (SAT) scores that began in the 1960s. The extent to which declining test scores are an important part of the productivity story is controversial. We have been told by different people whose judgment we respect that, on the one hand, this issue is a key one for the slowdown and, on the other hand, that the test score evidence is meaningless. It is worth taking a look.

In earlier work, Baily considered whether the decline in SAT scores could have been an important cause of the post-1973 growth decline.[24] He concluded that it could not have been, because the decline was not large enough, and the new entrants to the labor force with the lower scores did not make up a large enough fraction of the work force. A recent study by John Bishop has investigated not only SAT scores, but a variety of different measures of general intellectual achievement (GIA).[25] He considers one important set of tests that has been given to

23. Frank Levy, "Incomes, Families, and Living Standards," in Robert E. Litan, Robert Z. Lawrence, and Charles L. Schultze, eds., *American Living Standards: Threats and Challenges* (Brookings, 1988).

24. Martin N. Baily, "Productivity and the Services of Capital and Labor," *BPEA, 1:1981*, pp. 1–50.

25. John Bishop, "Is the Test Score Decline Responsible for the Productivity Growth Decline?" Working Paper 87–05 (Cornell University, January 6, 1988).

students in Iowa on a fairly uniform basis over many years, as well as others, for example those given by the armed forces. All the tests show that there has been a long-term trend of rising scores, which accelerated in the mid-1950s. The scores flattened out in the mid-1960s and declined in the 1970s.

Bishop uses a wage equation to estimate the impact of GIA on earnings in the cross section and then applies his coefficient estimate to determine the impact of the test score decline. His findings confirm that the decline in scores after 1967 did not cause the post-1973 slowdown. In fact, the upward burst of scores in the 1960s meant that average GIA for the work force was rising faster than trend during part of the slowdown period. Bishop does suggest, however, that the test score decline is now contributing to weak growth in the 1980s. He estimates that labor quality was reduced by 0.24 percent a year during 1980–87 as a result of the reduction in GIA.

While this figure is based upon his wage equation, Bishop obtains the 0.24 figure by scaling up his regression estimate with an adjustment for errors in variables. This scaling-up can be questioned. Certainly the regression estimate of the coefficient on GIA is likely to be biased downward relative to the true coefficient, because his proxy for GIA in the cross-sectional data is only a weak one. But estimating the impact of trends in GIA on labor quality using the scaled-up coefficient is correct only if the true trend in GIA is known. Since the time series trend in test scores is itself only a proxy, albeit a much better one, the adjustment to the regression coefficient may be too great.[26] The observed trends in test scores may reflect the emphasis the schools put on test score performance, rather than trends in underlying achievement. Schools stressed test scores in the post-Sputnik era and have started doing so again in recent years.

Overall, therefore, we accept the idea that GIA has grown less rapidly since 1973 than before, but a figure of 0.1 percent a year is a reasonable estimate of the magnitude of the decline in the quality of the work force

26. Daniel M. Koretz has recently reviewed the trends in test scores and their implications in a study for the Congressional Budget Office. He argues that the educational community is so diverse that the changing trends could not reflect anything different in what the schools were doing. We disagree. There certainly was a general change in social attitudes toward testing in the 1960s, but this shift then affected what happened in the schools. Congressional Budget Office, *Educational Achievement: Explanations and Implications of Recent Trends* (Government Printing Office, August 1987).

that is additional to the adjustments Denison makes for the quantity of education and other labor-force changes. Thus adding 0.1 point decline to Denison's 0.09 point decline in adjusted labor input relative to total hours, from table 8, would yield roughly a 0.2 point contribution of labor quality measurement to the post-1973 slowdown.

Issues in the Measurement of Output and Capital Input

The vast literature on output and input measurement is filled with disputes, some but not all of which have been cleared up in recent years.[27] We focus on the central measurement issues related to the productivity slowdown. At the most general level, there is an inevitable arbitrariness in the extent to which our measure of final output represents a broad measure of "welfare" or a narrower measure of currently produced physical output sold on the market.

Denison and others have recognized that no single generally acceptable measure of welfare can be constructed.[28] There is no straightforward way to measure the welfare cost of increased crime, congestion, and pollution of the air and water, nor the welfare benefit of improved medical care and of completely new products like the automobile, air conditioning, and home computers. And how are we to compare the present danger of nuclear war with past hazards, some of which are recalled by Denison in a memorable passage:

> Who would now think to consider the danger of attack by hostile Indians? Or the risk of being doused by slops thrown from windows as he walks the city streets? Even the very recent elimination of refrigerator doors that cannot be opened from within, and cost the lives of so many children, is almost forgotten. The annual series for "persons lynched" appears in the Census Bureau's *Historical Statistics* but not in its current *Statistical Abstract*.[29]

27. Basic references include Franklin M. Fisher and Karl Shell, *The Economic Theory of Price Indices* (New York and London: Academic Press, 1972); Panel to Review Productivity Statistics, *Measurement and Interpretation of Productivity* (Washington, D.C.: National Academy of Sciences, 1979); and Jack E. Triplett, "Concepts of Quality in Input and Output Price Measures: A Resolution of the User-Value Resource-Cost Debate," in Murray F. Foss, ed., *The U.S. National Income and Product Accounts: Selected Topics*, Studies in Income and Wealth, vol. 47 (University of Chicago Press for NBER, 1983), pp. 296–311.

28. Edward F. Denison, "Welfare Measurement and the GNP," *Survey of Current Business*, vol. 51 (January 1971), p. 13.

29. Ibid., p. 5.

Yet it goes too far to retreat entirely to a market-produced criterion for output. William Nordhaus, James Tobin, Richard and Nancy Ruggles, and Robert Eisner have produced estimates of some or all of the nonmarket activities that produce economic welfare. For our purposes, we need to have a conceptual framework for perspective on the evolution of the quality of marketed consumer services, some of which reflect new products made possible by the computer and other electronic capital goods, some of which represent substitution for formerly nonmarket activities produced in the home, and others of which increase "convenience." In this section we concentrate on issues related to quality improvements in durable goods and defer the discussion of convenience to our case study of retail trade.

COMPUTERS AND OTHER PRODUCER DURABLES

Durable goods are normally an input into the production of goods and services consumed by final users. Producer durables are an input, along with labor, structures, energy, and materials, in the production of consumer and producer goods. Consumer durables may also be considered an input, producing transportation services or household services. The crucial step in developing adequate deflators for durable goods, and hence in assessing the computer explosion, is to recognize that final market product (Q) is produced by a vector of market-purchased input characteristics (X):

$$Q = Q(X),\ Q_X > 0,\ Q_{XX} < 0. \tag{2}$$

An input characteristic is defined as any attribute of a market-purchased input that has a positive marginal product, including in the case of durable goods the horsepower and physical dimensions for a truck, or memory size and calculations per unit of time for a computer. In Triplett's more precise definition, a quantity is an input characteristic if it reduces the unexplained variation in output, given the explanation contributed by all the other arguments in the production function.

In determining the proper treatment of innovations in durable goods, we start with the types of shifts in the performance-to-price ratio that have been typical throughout the evolution of the computer industry, and then apply the same ideas to changes in energy efficiency or other aspects of user cost. We can call new-model introduction "proportional" when it raises the performance of a good by increasing its built-in quantity of characteristics (X) in the same proportions as the resources used by the supplying industry.[30] Such an innovation occurs when a new model is introduced that is larger or heavier and costs proportionally more to produce. In contrast, a nonproportional innovation raises performance by a greater percentage than the increase in resource cost.

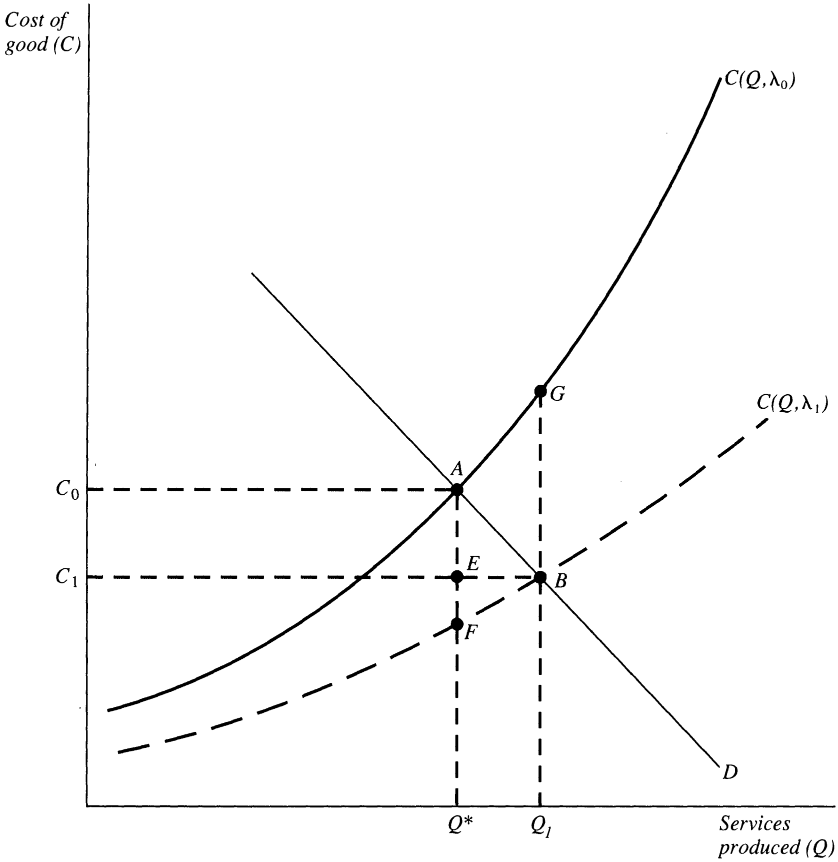
While nonproportional process innovations that improve the productivity of inputs in the manufacture of *given models* of durable goods occur continuously, the nonproportional innovations that concern us here are those involving both changes in processes and in product specifications that occur when a new model is introduced. Such an innovation takes the form of a downward shift in the real cost of producing a given quantity of characteristics, say computer calculations. There is no shift in the user firm's production function (equation 2), since a single calculation still produces the same amount of final output (Q). The quality change, though nonproportional, is not costless, since the reduction in cost must consume managerial and R&D resources, or else it would have occurred long ago.

We depict a nonproportional new-model introduction in figure 2 by plotting two upward-sloping lines plotting the cost function $C\ (Q,\lambda)$ of old and new models of a particular type of durable good corresponding to two different values of the technical shift parameter λ. Initially, output level Q^* is produced at an input cost of C_0 at point A. The technological change represented by the shift from λ_0 to λ_1 improves quality by raising the quantity of input characteristics relative to their cost. This raises the demand for characteristics and the level of output, depicted by Q_1 in the diagram. The unit cost of the durable good (C_1) could be either higher or lower than in the initial situation (C_0). In the diagram the unit cost declines along the demand curve D, but the unit cost could increase if the demand curve were to shift to the right.

The change in the input price index is simply $-\ AF$, the vertical downward shift in the supply schedule itself. This would be measured in practice by taking the observed reduction in the price of the machine ($-AE$) and adding an adjustment factor equal to the change in output produced by the extra characteristics (EF) times the marginal cost (EF/EB) of building the extra input characteristics. This extra adjustment

30. This discussion summarizes part of chapter 2 of Robert J. Gordon, *The Measurement of Durable Goods Prices* (University of Chicago Press for NBER, forthcoming).

Figure 2. Technological Change and Declining Costs



is what is accomplished by estimating regression coefficients for the value of greater speed and memory in the case of the hedonic price index for computers used now in the NIPA. For instance, if a new model computer costs 10 percent less than an old model, but the regression coefficients imply that its greater speed and memory represent 40 percent higher quality, the measured price decrease is not 10 percent but rather 50 percent. In the example of the hedonic price index for computers now used in the NIPA, such calculations lead to an annual rate of price decline of about 20 percent. Shifts in the supply curve like that depicted in figure 2 have greatly increased the performance of mainframe computers without reducing their average price. For instance, the IBM model 4381-2, introduced in 1984, cost the same $0.5 million as the IBM model 7070 introduced in 1960, but had a calculating power 1452 times as great.[31]

The idea of nonproportional quality change can be extended to changes in energy efficiency and other characteristics that affect user cost. Now a nonproportional quality change is one that raises the "net revenue" earned by a machine (gross revenue minus operating cost, prior to the deduction of depreciation and interest) relative to the machine's cost, holding constant the price of output, energy, and labor when comparing the net revenue of two alternative models incorporating different technologies. To highlight the meaning of nonproportional in this case, consider the introduction of a more expensive new refrigerator model. If it saves energy with existing technology, and if the extra cost of the refrigerator is equal to the present value of energy saving, then this is a proportional change and just moves us along a fixed cost function, as between *A* and *G* in figure 2. But the invention of a new form of insulation that allows all refrigerators to be more efficient at the same refrigerator price would represent a nonproportional change shifting down the supply curve in figure 2, and this would call for a quality adjustment and a decline in a properly measured price index. The most dramatic example of such a change was the invention of the jet plane, which raised the net-revenue-generating ability of a DC-8 compared with a DC-7 by a factor of 10 at a capital cost only three times higher.

The ultimate test of this approach to the measurement of durable goods prices is to compare its predictions with the verdict of the used asset market, as has been done with used aircraft prices and could now be done with the prices of used PCs.[32] How does the approach differ from current NIPA deflation procedures? The NIPA computer deflator treats nonproportional quality change by the method that we recommend, so that there is no dispute in principle. But many improvements

31. This example is taken from Robert J. Gordon, "The Postwar Evolution of Computer Prices," in Dale W. Jorgenson and Ralph Landau, eds., *Technology and Capital Formation* (MIT Press, forthcoming), table 3.9. Over the same interval, 1961–83, the average unit price of a mainframe computer increased from $0.3 to $1.0 million (table 3.1).

32. On used aircraft, see Gordon, *Measurement of Durable Goods Prices*, chap. 4. On used PC prices, see Mark Lewyn, "Here's What to Look for in Used PCs," *USA Today*, August 17, 1988.

in the ratio of performance to price for durable goods are missed, either because official price indexes (mainly components of the producer price index) make inadequate allowances for quality change, or because price indexes are simply missing for important types of capital equipment, such as electronic telephone switching equipment, where technological progress has been rapid. As for improvements in energy efficiency, there are no explicit adjustments in the NIPA.[33]

IMPLICATIONS OF NEW MEASURES OF DURABLE GOOD DEFLATORS

In a project to create alternative price indexes for durable goods, Gordon has combined more than 25,000 annual price observations from sources independent of the PPI and CPI, including the Sears catalog, *Consumer Reports* price quotes and quality evaluations of eight types of appliances and seven other products, used auto and tractor price manuals, government regulatory data on the price and performance of commercial aircraft and electric generating stations, independent data on computer prices, and American Telephone and Telegraph data on the price and performance of telephone transmissions and switching equipment.[34] For many of the products the study carries out the conventional BLS methodology by comparing only identical models in pairs of successive years over the full period 1947–83. For products where operating cost data are available for particular models, adjustments are made for energy efficiency (aircraft, electric generating equipment, railroad locomotives, consumer appliances) and in one case (TV sets) for repair frequency.

The results of the study yield radical conclusions for some issues, while making only a slight contribution to an understanding of the productivity slowdown. Some of the implications for growth rates of producers' and consumers' durable goods are summarized in table 9.

33. Improving fuel economy for automobiles has been implicitly taken into account, at least in part, by the decision of the BLS not to treat the shrinking dimensions of "downsized" automobiles as a decline in quality. If automobiles of a given size and performance have better fuel economy now than 20 years ago, however, this aspect of quality improvement has not been taken into account.

34. The latter collected for a study by Kenneth Flamm, "Economic Dimensions of Technological Advance in Communications: A Comparison with Computers" (Brookings, March 1988).

Table 9. Alternative and NIPA Deflators and Investment-to-GNP Ratios for Durable Goods, 1947–83, Selected Years

Item	*1947*	*1960*	*1973*	*1983*
1. PDE deflator (1982 = 100)				
Alternative[a]	59.60	60.10	58.40	99.10
NIPA	20.60	35.70	47.30	99.50
Alternative-NIPA ratio	2.89	1.68	1.23	1.00
2. PDE/GNP (1982 dollars, percent)				
Alternative[a]	2.58	3.07	6.02	7.16
NIPA	6.96	5.00	7.27	7.14
Alternative-NIPA ratio	0.37	0.61	0.83	1.00
3. Consumer durable expenditures/GNP (1982 dollars, percent)				
Alternative[a]	3.27	4.66	7.38	8.64
NIPA	5.30	5.89	8.03	8.65
Alternative-NIPA ratio	0.62	0.79	0.92	1.00
4. Total durables/GNP (1982 dollars, percent)				
Alternative[a]	5.85	7.73	13.40	15.80
NIPA	12.26	10.89	15.30	15.79
Alternative-NIPA ratio	0.47	0.71	0.88	1.00

Source: For further discussion of the calculation of alternative deflators, see Robert J. Gordon, *The Measurement of Durable Goods Prices* (University of Chicago Press for NBER, forthcoming), especially tables 12.4, 12.5, and 12.10.

a. Alternative based on new detailed price data assembled by Gordon, using Törnqvist indexes that weight the annual percentage change in components of real output in each subcategory between years t and $t+1$ by the average of the nominal value weights in the two adjacent years.

The new producers' durable equipment (PDE) deflator rises 3.0 percent a year *more slowly* than the NIPA PDE deflator for the full 1947–83 period, with a somewhat larger negative "drift" in the first half. Where applicable, the same data are reweighted to create a new consumer durable deflator. The most startling change in the numbers is for the ratio of real PDE spending to real GNP, as shown in the second section of table 9. There is a smaller difference for the consumer-durables-to-GNP ratio, but still a major revision in the ratio of total durables spending to GNP, which rises between 1947 and 1983 by 29 percent in the NIPA version and 170 percent in the new version.

The finding on line 2 of table 9 that the ratio of equipment investment to output has increased rapidly during the postwar period, instead of remaining roughly constant, has important implications for longstanding debates regarding the interpretation of growth theory and the sources of growth. The new data imply that the growth process has been characterized by more rapid growth in real investment than in real output. This carries over to more rapid growth in real capital input than in real output, and to a steady increase in the capital-output ratio throughout the postwar period. In contrast, the steady state in a standard neoclassical growth model describes a situation in which investment, capital, and output all grow at the same rate, and in which the investment-output and capital-output ratios are constant. The mechanism by which these ratios grow continuously in the new data is not a steady increase in the share of saving in total income, but rather a steady shift in relative prices that cheapens capital equipment relative to other types of output.

Soon after his original articles on growth theory and measurement, Robert Solow advanced the "embodiment hypothesis" that productivity gains result, in large part, from the installation of capital goods that embody new technologies. Some studies, most recently that of Angus Maddison, have interpreted this hypothesis to imply that explicit adjustments should be made to the BEA measures of the capital stock to account for embodied quality improvements.[35] Presumably, the embodiment hypothesis implies that better data on quality improvements in capital goods would substantially reduce the growth accounting "residual," that is, the growth rate of MFP, by raising the growth rate of effective capital input.

The new price deflators for capital equipment used in table 9 achieve Maddison's desired adjustments for performance improvements in successive vintages of capital goods and thus can be used both to assess Solow's embodiment hypothesis and to determine whether the improved data substantially reduce the growth rate of MFP. As shown in table 10, the implications of the new data for MFP growth are surprisingly small and in this sense serve to refute Solow's embodiment hypothesis that a large fraction of MFP growth is attributable to unmeasured improvements in capital quality. The bottom section of table 10 shows that the

35. Solow's original investigation of the sources of economic growth was "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, vol. 39 (August 1957), pp. 312–20. His embodiment hypothesis was set forth in "Investment and Technical Progress," in K. J. Arrow, S. Karlin, and P. Suppes, eds., *Mathematical Methods in the Social Sciences* (Stanford University Press, 1959). For Maddison's analysis, see "Growth and Slowdown in Advanced Capitalist Economies: Techniques of Quantitative Assessment," *Journal of Economic Literature*, vol. 25 (June 1987), pp. 649–98, esp. pp. 662–64.

Table 10. Effect of Alternative Durable Goods Deflators in Sources-of-Growth Calculation, 1947–83, Selected Periods

Annual percentage growth rate over interval, except where noted

Item	*1947–60*	*1960–73*	*1947–73*	*1973–83*	*1947–83*
Private GNP					
Alternative[a]	3.68	4.14	3.91	2.08	3.40
NIPA	3.35	4.02	3.68	1.82	3.17
Alternative-NIPA ratio	0.33	0.12	0.23	0.26	0.23
Capital input					
Alternative[a]	4.60	5.73	5.17	4.97	5.11
NIPA	3.10	3.87	3.49	3.56	3.51
Alternative-NIPA ratio	1.50	1.86	1.68	1.41	1.60
Capital contribution[b]					
Alternative[a]	1.15	1.43	1.29	1.24	1.28
NIPA	0.78	0.97	0.87	0.89	0.88
Alternative-NIPA ratio	0.37	0.46	0.42	0.35	0.40
Private business labor hours	0.79	1.93	1.36	1.00	1.26
Labor contribution[c]	0.59	1.45	1.02	0.75	0.95
Multifactor productivity[d]					
Alternative[a]	1.94	1.26	1.60	0.09	1.17
NIPA	1.98	1.60	1.79	0.18	1.34
Alternative-NIPA ratio	−0.04	−0.34	−0.19	−0.09	−0.17

Sources: Gordon, *Measurement of Durable Goods Prices*, table 12.11. Original data for private business labor hours are from NIPA, table 6.11, extrapolated back from 1948 to 1947 by use of full-time equivalent employment from NIPA, table 6.7A.

a. See table 9, note a.
b. Equals capital input times 0.25.
c. Equals labor input times 0.75.
d. Equals growth in output minus capital contribution minus labor contribution.

alternative deflators reduce MFP growth by 0.19 point for 1947–73 and 0.09 point for 1973–83. Thus improved capital quality explains just one-tenth of the 2.0 growth rate of MFP registered in table 1 for the pre-1973 period, and about one-quarter of the 0.4 growth rate of MFP from table 1 for the post-1973 period. As for the MFP slowdown itself, the new capital data explain 0.10 point (that is, the difference between 0.19 and 0.09). There is a simple reason why the radical revisions to capital goods deflators have such small implications for MFP growth: the new deflators make not just capital input but also output grow faster, so that the impact on MFP growth is much less than on output or capital separately.

DENISON ON THE COMPUTER DEFLATOR

The approach to capital good prices just described has been rapidly gaining adherents, but not everyone is persuaded. Edward Denison has been particularly critical of the new hedonic price deflator for computers now used in the NIPA. Denison's opposition to the computer price index derives from his belief that "the line between the contributions of capital and advances in knowledge, in particular, should be so drawn that the former measures growth that results from saving and investment . . . and the latter measures comprehensively growth that results from advances in knowledge that permit goods and services to be produced with less input."[36]

Denison's preferred method would purge from capital input growth all contributions of advances in knowledge, both present and past. The output of capital goods would not be allowed to reflect the improvements in the ratio of performance to price that were dubbed nonproportional in the previous section. But even further, the effects of process innovations within the capital-goods-producing industries would also be excluded, for example, improvements that allow IBM to reduce the labor input in making a given model computer. If this procedure were followed in full, every increase in the ratio of capital goods output to inputs would be excluded, implying that capital goods output would be measured by labor input. His method would convert a 120 percent increase in the output of the nonelectrical machinery industry over 1973–86 into a 2 percent decline and set productivity change in that industry at zero by definition.[37]

While Denison's desire to track down all contributions of advances in knowledge at the aggregate level is a worthy one, it makes less sense at the industry level. For many purposes, such as computing the private and social return of research and development, we want to classify innovations in the industry where they occur. Just as the NIPA computer deflator credits the nonelectrical machinery industry with the output achieved by research in the computer industry, so our preferred approach would apply the same principle uniformly, crediting the airframe and aircraft engine industries rather than the airlines for the invention of the jet plane, and the electrical machinery industry rather than the electric utilities for the radical improvements in electric generating equipment that occurred up through the late 1960s.

Denison objects to the NIPA computer deflator not just on principle, but also because it introduces inconsistency into the national accounts.

36. Denison, "Estimates of Productivity Change," p. 39.

37. NIPA, table 6.2, line 20, and table 6.7B (full-time equivalent employees), line 20.

Computers are compared by a marginal product criterion while other products are treated as equivalent if they would have cost the same to produce in the base period. Here Denison is partly correct, for indeed the NIPA deflators take no account of radical improvements in quality in some other types of durable goods. But rather than move backward toward the elimination of the computer deflator, we would move forward by applying consistent principles to these other products, as has been carried out for several of the most technologically progressive producers' durables in the results summarized in tables 9 and 10.

COMPUTER PRICES AND THE USE OF CONSTANT-DOLLAR INDEXES

The NIPA measure real output, investment, and capital in units of 1982 dollars. The autos or loaves of bread or computers produced in 1987 are valued in real GNP based upon the price paid for these items in 1982. Even if we accept that the price indexes that are used do a good job of adjusting for changes in the nature of autos or bread or computers over time, this procedure is still flawed because in fact relative prices have changed. The marginal utility of consumer goods was not the same in 1987 as in 1982. Nor was the marginal product of investment goods. The effect of changing relative prices is not random. Goods where technological progress has been rapid have falling relative prices and increasing sales volumes. The use of base-period prices overweights the growth of these dynamic commodities in years following the base year and underweights them in years preceding the base year.

In the case of the computer, the distortion of real output created by the declining relative price is likely to be important because of the rapidity of the decline. In 1969, investment in office computing and accounting machines (OCAM) was 6.1 percent of total nonresidential equipment investment in current dollars, but only 1.6 percent in constant 1982 dollars. In 1986, OCAM expenditures were 11.3 percent of equipment spending in current dollars, but were 22.9 percent in constant dollars. Constant-dollar base-weighted investment series imply that the computer industry disappears as you go back a few years, and it explodes and takes over the total as you go forward in time.

Some of the opposition to the new computer price index may stem from the observation that when someone purchased a computer in 1975 that had the same capabilities as today's PC, this 1975 computer was used intensively and was essential to the tasks being performed. Today, many computers sit idle in peoples' offices. However, the way computers are used today is what one would expect, given their low price. The way to deal with the problem of the declining marginal productivity of computers is to use Törnqvist current-year nominal shares in constructing aggregate series. How much difference would this make?

Gordon's work provides one example of the radical effects that occur when Törnqvist indexes are used in place of the NIPA fixed base-year method. His basic results reported in table 9 are based on Törnqvist indexes that weight the annual percentage change in components of real output between successive years t and $t+1$ in each subcategory by the average of the nominal value weights in the two adjacent years. This technique has the effect of weighting each computer calculation by its steadily falling price, corresponding to its steadily declining marginal product for the user. To show the effects of the NIPA fixed-base-year method, Gordon calculates the PDE deflator in 1972 prices implied by the existing NIPA deflators for the 22 subcategories of PDE. The results are striking and suggest what would have occurred if the BEA had introduced its computer deflator 10 years earlier but kept its existing fixed-base-year methodology for aggregation. Instead of roughly doubling from 1972 to 1983, as occurs with the Törnqvist weights, the NIPA implicit deflator would have increased by only 30 percent and in 1983 would have been at the same level as in 1978. The weight of OCAM in total PDE in 1983 would have been 60 percent, in contrast to its 11 percent nominal share.

To assess the effects to date of the use of fixed 1982 price weights, we have roughly separated computers from the rest of investment and then combined the two parts using current-year shares (see table 11). The results are shown first for OCAM and PDE investment. It is clear that computers are driving the dynamism of equipment investment since 1979. Without computers, PDE investment was actually falling during 1979–87. The distortion caused by the underweighting before 1982 and the overweighting afterwards makes a difference, raising real PDE growth 0.52 percent a year during 1973–79 and lowering it 0.44 percent a year during 1979–87.

We extrapolated recent investment trends forward and found that by 1996 the constant-dollar base-weighted measure of PDE would be growing three times as fast as the real output measure calculated with current-year weights. Presumably the NIPA's procedures or the base year or both will be adjusted by then, but these results provide a warning

Table 11. Alternative Measures of Real Output Growth for Producers' Durable Equipment and Office Computing and Accounting Machines, 1973–79 and 1979–87
Percent per year

Item	*1973–79*	*1979–87*
Conventional		
NIPA PDE	4.33	2.64
NIPA OCAM	19.79	24.28
NIPA PDE less OCAM	3.85	−0.37
Alternative		
PDE with current-year share weights for OCAM[a]	4.85	2.20

Sources: PDE and OCAM for 1973 and 1979 taken from NIPA, table 5.7. PDE and OCAM for 1987 taken from *Survey of Current Business*, vol. 68 (July 1988), table 5.7.
a. The rate of growth between two years t and $t-1$ is calculated as the weighted average of the rates of growth of OCAM and PDE excluding OCAM, both series measured in 1982 dollars. The weights are the average current-dollar shares of OCAM and PDE excluding OCAM in total current-dollar PDE, where the average is over the two years t and $t-1$.

about the validity of the investment series now being released and those coming in the next few years. Already these procedures are causing a significant distortion of official data on growth in both PDE and GNP. Although the BEA now publishes chain-linked deflators for PDE and GNP that approximate the use of Törnqvist indexes, it does not use these deflators to compute real PDE or GNP. Superior measures of real PDE and GNP growth can be obtained by subtracting the difference between the chain and implicit deflators from the existing official estimates of growth in real PDE and GNP. The resulting calculation shows that real PDE growth was overstated by 3.2 percent and real GNP growth by 0.8 percent for the four quarters ending in 1988:2.[38] Policy-makers may be led by official data on real growth to overstate the pace of the current economic expansion.

Computer Power in the Using Industries

Whatever one's view of particular measurement procedures, there is no question that the computer has been enthusiastically adopted by

38. The existing real GNP series is equal by definition to nominal GNP divided by the implicit GNP deflator. Our preferred real GNP series is equal to the same nominal GNP value divided by the BEA's chain deflator for GNP, and the same for our preferred real PDE series. Growth rates of NIPA implicit and chain-weighted deflators are from *Survey of Current Business*, vol. 68 (July 1988), table 8.1.

American industry and has brought about major changes in the way business is conducted. A fundamental paradox in U.S. productivity over the past 20 years is that during a period seen by many as one of rapid or even accelerating technological change, productivity growth has been weak. Moreover, the sectors in which computers and other electronic equipment are being used are showing particularly slow growth. ALP in these industries should certainly have benefited from the electronics revolution. Why not?

As a first step toward answering this question it is worth looking at electronics investment. Table 12 shows the net stocks of computers, communications equipment, and related capital by industry for 1987 and earlier periods. These data are based upon industry of ownership, which is a problem for this type of capital, for which leasing is important. In addition, the coverage is much more extensive than just computers. Communications equipment is clearly an important part of the total.

The manufacturing sector is not a big owner of the electronic equipment it produces. In fact, within manufacturing, the machinery industry is the largest owner of electronic equipment, where presumably its use and ownership are tied quite closely to its production. Outside of manufacturing, the communications industry stands out as a service industry that has invested heavily in electronics and achieved rapid productivity growth by doing so. Communications is the exception that proves the rule, however, since the other white-collar areas—trade, finance, insurance, and real estate, and services—are all fairly intensive in their shares of electronics capital, and all have had weak growth.

Analysts have offered five reasons why electronics investment has not paid off in greater productivity growth in the white-collar service industries.[39]

First, dramatic changes in technology can make productivity worse before it gets better. People have to be retrained, and companies have to learn how to use the new technology efficiently. This hypothesis

39. Baily and Chakrabarti, *Innovation*, pp. 86–102; Office of Technology Assessment, *Automation of America's Offices, 1985–2000* (GPO, 1985); Stephen S. Roach, "America's Technology Dilemma: A Profile of the Information Economy," Special Economic Study (Morgan Stanley, April 22, 1987); H. Allan Hunt and Timothy L. Hunt, *Clerical Employment and Technological Change* (Kalamazoo, Michigan: W. E. Upjohn Institute, 1986); Gary W. Loveman, "The Productivity of Information Technology Capital: An Econometric Analysis" (Massachusetts Institute of Technology, January 31, 1986); and Paul Osterman, "The Impact of Computers on the Employment of Clerks and Managers," *Industrial and Labor Relations Review*, vol. 39 (January 1986), pp. 175–86.

Table 12. Net Stocks of Computers, Office and Accounting Machinery, Communications Equipment, Instruments, Photocopiers, and Related Equipment, by Industry, Selected Periods, 1960–87

Billions of 1982 dollars except as noted

Industry	*Total nonresidential capital, 1987*	*Computers and communications equipment, 1987*	*Computers and communications equipment as a percent of total, 1987*	*Computers and communications equipment as a percent of total*	
				1960–69	*1970–79*
Manufacturing	763.30	77.25	10.1	1.6	2.8
Nonmanufacturing	2,810.00	454.07	16.2	4.4	6.7
Mining	256.80	0.18	0.1	0.1	0.1
Construction	50.10	2.15	4.3	0.7	0.5
Transportation	254.34	2.88	1.1	0.5	0.6
Rail	96.18	0.50	0.5	0.5	0.7
Nonrail	158.16	2.38	1.5	0.7	0.5
Air	36.69	0.86	2.3	0.7	0.4
Trucking	48.20	0.17	0.4	0.5	0.1
Other	73.27	1.34	1.8	0.8	0.8
Communications	317.66	172.73	54.4	30.6	40.8
Public utilities	448.86	19.74	4.4	0.6	1.1
Total trade	413.48	52.82	12.8	0.9	2.5
Retail	237.59	9.02	3.8	0.5	1.1
Wholesale	175.89	43.80	24.9	1.8	4.9
FIRE	719.78	143.94	20.0	5.7	6.0
Finance and insurance	234.73	90.51	38.6	3.5	4.7
Banks	109.15	37.17	34.1	1.9	3.9
Credit agencies	68.61	18.19	26.5	3.1	3.3
Securities	6.61	3.88	58.7	3.9	8.3
Insurance carriers	47.24	30.12	63.8	4.4	7.2
Insurance agents	3.12	1.16	37.2	15.6	12.0
Holding companies	16.90	9.10	53.8	8.0	10.2
Real estate	468.15	44.32	9.5	6.1	6.3
Services	348.98	59.63	17.1	6.5	8.3
Hotels	61.34	1.36	2.2	0.1	0.2
Personal	13.00	1.79	13.8	9.3	13.6
Business	92.04	25.68	27.9	7.9	10.9
Auto repair	60.34	1.01	1.7	0.5	0.2
Miscellaneous repair	7.53	0.28	3.7	0.4	0.5
Motion pictures	6.32	2.97	47.0	15.0	31.5
Amusement	21.65	4.65	21.5	9.7	12.3
Other	81.76	21.89	26.8	12.2	13.3
Health	52.37	16.39	31.3	16.0	19.2
Legal	7.06	1.32	18.7	2.9	4.0
Educational	2.06	1.19	57.8	9.5	12.2
Other	20.27	2.98	14.7	10.0	6.6

Sources: 1987 total nonresidential capital for manufacturing, mining, and construction taken from "Fixed Reproducible Tangible Wealth in the United States," *Survey of Current Business*, vol. 68 (August 1988), table 4. Remaining data from Stephen S. Roach, "America's Technology Dilemma: A Profile of the Information Economy," Special Economic Study (Morgan Stanley, April 22, 1987), for 1960–69 and 1970–79; and Stephen S. Roach, "White-Collar Productivity: A Glimmer of Hope?" Special Economic Study (Morgan Stanley, September 16, 1988), for 1987.

suggests that future productivity gains will be rapid, as the training and learning take place. That productivity growth has been decelerating in nonfarm nonmanufacturing, however, bodes poorly for this hypothesis.

Second, the electronics revolution involves the acquisition, processing, and distribution of information. This information may be used to take customers, profits, or capital gains away from other companies. If computers are used extensively in ways that redistribute wealth rather than increase it, then their productivity effects will be reduced. For example, some computers and telephones are used in marketing activities that are largely forms of advertising.

Third, the technology of computers may encourage waste and inefficiency. Computers provide a flow of services to companies that the companies do not know how to value. White-collar groups sometimes measure their performance on the basis of the amount of information or paperwork they generate rather than on its value to a company.

Fourth, computers may improve working conditions. Does any secretary lament the passing of purple ditto masters on which corrections were made with razor blades? Does not the mastery of soft fonts and page layout provide more job satisfaction than drone-like repetitive retyping of successive drafts? This view would argue that the stagnation of real wages in some occupations is exaggerated, because wage payments have been held down as a compensating differential for improved job satisfaction. We firmly believe that this effect constitutes part of the overall impact of computers but do not pursue it for lack of hard evidence.

The final point, and the one most directly related to this paper, is that computers and related equipment may be providing valuable services to customers that are not being picked up in the official output data. This point provides the wedge through which we may attempt to locate errors in the measurement of consumer service output by arguing that deflators for consumer expenditures on services do not adequately capture quality improvements created by computers.

In future work we plan to explore several of these alternative hypotheses more fully, but in this paper we are concentrating on the idea that computers have made it harder to measure output. First, we look at general examples of ways in which computers are increasing convenience or providing other services. Then, in our first case study, we look more closely at finance, insurance, and real estate, a sector that includes many industries that are heavy users of computers and related equipment. Just the beginning of a list of computer-based consumer service improvements and innovations would include the following.

—The transportation sector offers pre-assigned seats and boarding passes, "no-stop" check-in, frequent-flyer plans that amount to unmeasured price reductions, and price discrimination that has reduced prices for consumers relative to businessmen. Consumers can delay their Christmas shopping because of overnight delivery services that offer continuous package tracking for the rare instances when something does not arrive as promised.

—The retail trade sector offers better inventory control, fewer stockouts, and most notably the radical increase in the variety of items carried that we document in our case study below. Drugstore chains have introduced computerized prescription records that allow prescriptions to be refilled from any branch store in the chain.

—The finance sector offers all-in-one cash management accounts, costless portfolio diversification for even the smallest investors through no-load mutual funds, automatic telephone machines allowing almost instantaneous credit card approvals, fast bill-paying by phone or PC, and 24-hour money machines.

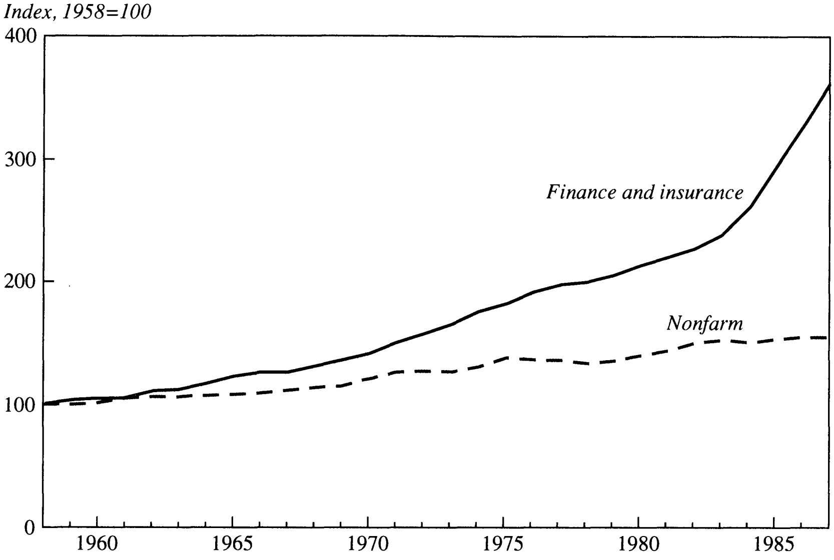
—Hotel chains provide frequent-stayer services that, upon recitation of a single number, allow a reservation to be made without the need for a telephone caller to mention a name, credit card number, or room preference. The hotels themselves provide pre-printed registration forms, no-stop check out, and clerks who answer the room-service phone by telling the guest his name and room number, instead of vice versa.

—Restaurants, supermarkets, and, less happily, hospitals provide itemized bills for those curious to know where their money has gone.

Case Studies of Measurement in Specific Sectors

There is only so much to be learned from aggregative data about the potential size of measurement errors. To go further it is essential to be specific about how price deflators and real output are calculated in particular cases. Thus we turn now to a case study approach, choosing examples where we suspect from the outset that data errors may be important. We hope also that these case study examples will suggest ways of improving output measurement that could be used in other sectors.

Figure 3. Gross Capital Stock per Employee (1982 Dollars), Finance and Insurance and Total Nonfarm, 1958–87



Finance, Insurance, and Real Estate

The financial services industry is extremely dynamic, with rapid employment growth, a high rate of capital accumulation, and a steady flow of new products.[40] The insurance industry seems ideally suited to the data-handling capabilities of modern computers; both it and the finance industry have invested heavily in new capital. Figure 3 shows that capital per worker in this industry has grown much more rapidly than in the economy as a whole. In this case study we look first at the finance, insurance, and real estate sector as a whole, and review problems in the measurement of output in the insurance and real estate parts of the sector. We then focus on financial services and the banking industry in particular as a case study within the case study. Banking is an important part of the total and is an industry where data availability allows some alternative productivity estimates to be made.

Table 13. Shares of Industries in Finance, Insurance, and Real Estate Output, 1948–87, Selected Years
Percent

Industry	1948	1965	1973	1979	1987
Banking	22.51	18.79	19.56	19.43	18.43
Credit agencies	1.61	2.05	1.73	1.83	2.42
Securities	3.95	4.04	3.91	3.37	6.21
Insurance carriers	12.72	11.85	11.10	10.25	10.79
Insurance agents	8.63	7.94	7.09	6.42	6.97
Real estate[a]	49.56	54.20	54.78	56.49	52.00
Holding companies	1.02	1.13	1.82	2.22	3.18

Source: *Survey of Current Business*, vol. 68 (July 1988), tables 6.2 and 1.24.
a. Real estate less gross owner-occupied housing and farm housing.

The finance, insurance, and real estate sector includes banking, credit agencies, the securities industry, insurance carriers, insurance agents, real estate, and holding and other investment companies. The real estate industry includes real estate agents and rental property. The shares of the components of the overall sector over the postwar years are shown in table 13. In 1987 real estate alone accounted for over half of the sector in terms of GDP originating and for almost 23 percent of employment. The imputed income of owner-occupied houses (and farm dwellings) is excluded from the business sector and from table 13.

Table 14 provides the labor productivity growth rates for each of the industries in the sector and for the sector as a whole. The labor input used in the calculations in this table is the "number of persons engaged in production," including full-time equivalent employees and those who are self-employed. The figures for the sector as a whole, therefore, differ somewhat from those given in table 4, which used the BLS estimates of total hours. The BEA uses labor input to extrapolate real output changes for the banking, credit, securities, and holding company sectors, and it should, therefore, show no productivity growth except for mix effects within the industries. That is roughly confirmed by table 14. Within the insurance and real estate industries, rents, premiums, and commissions are used to estimate nominal dollar output, and various deflators, to calculate constant-dollar output. Thus it is possible for the official data in these industries to show productivity growth. Indeed, the real estate sector showed substantial growth during 1948–65, and insurance carriers have achieved some productivity growth both pre- and post-1973, even though their performance was rather erratic over shorter periods.

Table 14. Labor Productivity Growth of Finance, Insurance, and Real Estate Sectors, 1948–87, Selected Periods[a]

Percent per year

Sector	*1948–65*	*1965–73*	*1973–79*	*1979–87*	*1948–73*	*1973–87*	*Change, 1948–73 to 1973–87*
Total	1.70	0.78	0.09	−0.79	1.41	−0.41	−1.82
Banking	0.00	0.02	0.01	0.08	0.00	0.05	0.05
Credit agencies	−0.21	−0.20	0.12	0.01	−0.20	0.05	0.25
Securities	0.08	−0.11	0.03	0.02	0.02	0.02	0.00
Insurance carriers	0.96	1.82	0.00	1.48	1.23	0.84	−0.39
Insurance agents	0.42	−0.06	−2.60	−0.42	0.26	−1.36	−1.62
Real estate[b]	4.84	−0.04	−0.05	−1.28	3.25	−0.76	−4.01
Holding companies	−0.36	−0.01	−0.12	0.03	−0.25	−0.04	0.21

Source: *Survey of Current Business*, vol. 68 (July 1988), tables 6.2, 1.24, and 6.10B.

a. Labor input measured as number of persons engaged in production including full-time equivalent employees and self-employed.

b. Real estate less gross owner-occupied housing and farm housing.

DEFLATORS IN REAL ESTATE AND INSURANCE

Because real estate and insurance made up 70 percent of total sector output in 1987, the quality of the deflators used in these industries is crucial to the accurate measurement of output in the sector as a whole. In practice the deflators used are either inappropriate or subject to substantial error.

The deflators used for the insurance industry are those developed for the industries being covered by the insurance. The auto repair cost index is used for auto insurance, medical costs for medical insurance, and so on. This deflator is applied to the nominal output of the industry, which is calculated as the difference between insurance premiums paid and claim reimbursements returned. Thus, the productivity weakness in the insurance sector is being driven by the escalation of cost indexes in the medical care area and in repair services, even though the insurance industry is engaged in an entirely different productive activity.[41] An example can illustrate how the problem distorts measures of insurance productivity. Suppose a medical insurance policy in the base year cost \$100. This policy, on average, paid for nine visits to a doctor at \$10 a visit. The remaining \$10 was retained by the insurance company to cover

41. Even if one accepted this approach to deflation there are serious problems with the particular deflators used. The medical care cost index neglects the tremendous quality changes that have taken place in this area.

Table 15. Rates of Growth of Price Deflator for Insurance and GNP, 1948–87, Selected Periods

Percent per year

Period	*Insurance carriers*	*Insurance agents*	*GNP*
1948–73	3.89	5.03	2.97
1973–79	9.61	9.13	7.69
1979–87	6.28	6.39	5.05

Sources: Data for 1948, 1973, and 1979 from NIPA, tables 6.1, 6.2, and 7.4. Data for 1987 from *Survey of Current Business*, vol. 68 (July 1988), tables 6.1, 6.2, and 7.4.

its costs and profits. Suppose that some years later an identical policy that cost $465 paid for nine visits at $50 a visit, with the insurance company retaining $15 and performing the same services as before. The BEA says that visits have risen in cost by a factor of five and computes a total real output of $93, of which only $3 consists of real gross product originating in the insurance industry. On this calculation, real output in insurance is way down, but, by assumption, true real output has remained constant. After excluding the insurance provided as an intermediate good to businesses, this distortion of real output does lead to a distortion of final GNP.

Obviously any distortion of real output in practice depends upon how the deflators actually used compare with true deflators, and these are unknown. The trends in the actual deflators are shown in table 15, together with trends in the implicit deflator for GNP as a benchmark. Given that the insurance industry has been able to benefit not only from computerization, but also from the widespread use of group policies, it is implausible that insurance costs should have risen faster than the price level for GNP.

An appropriate real output series for the insurance industry should be based upon the number of policies issued, with allowance made for changes in quality resulting from changes in the extent of coverage provided. An additional activity by life insurance companies, namely, the management of saving and mutual fund investment, should be treated comparably to that of other financial intermediaries.

The real estate industry's output is the rental income it receives and the commissions of realtors. This nominal output is deflated using rental cost indexes for residential and commercial rents. The series for commercial rents relies on reports in the industry of rental prices per square foot in major markets. Standard rental cost indexes have moved up rapidly since 1973, resulting in negative reported productivity growth in the industry. The problem in using the available rent indexes as deflators is that they do not adjust for changes in the quality of the property being rented. A commercial office building today is very different from one built 20 years ago. The rent index carries over some of the problems of the construction price index. It would be possible, at least in principle, to apply the hedonic technique using office space of different vintage to control for this quality change.

One important reason why one would expect "productivity" to have declined in the real estate sector is that the output in the industry is made up largely of the return to capital. Decisions to hold real estate property are based not only on its rental income return, but also on the expected capital gain from the property and the land it occupies. People are sometimes willing to invest in property with a low or even negative rate of return from rents, counting on a continuation of large capital gains. Market equilibrium in the 1970s and 1980s has resulted in a low marginal product for real estate investment.

FINANCIAL SERVICES

The output of the financial services industry is measured on the basis of labor input and thus ignores any output per hour gains by definition.[42] Clearly, alternative output series are needed, and we report now a fragmentary attempt to measure output per employee in the securities industry. Data for selected years on the number of shares traded in all markets, the market value of shares, and the number of persons engaged in the securities industries are shown in table 16.

42. Timothy F. Bresnahan has constructed a series for capital input for this industry using the new computer price series and then estimating the welfare gains from computerization. His work is of great interest, but does not provide an alternative measure of output for productivity analysis, nor is that his intention. Welfare measures are not the same as productivity measures. His procedure implies that output in the industry should be measured from labor and capital inputs, not just labor alone. Using both factors rather than one is certainly an improvement and would allow for labor productivity growth, but it does not provide the independent output series needed for productivity analysis. See Timothy F. Bresnahan, "Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services," *American Economic Review*, vol. 76 (September 1986), pp. 742–55.

Table 16. Number and Value of Shares Traded and Number of Employees in the Securities Industry, 1961–87, Selected Years

Year	*Number of shares (millions)*	*Market value of shares (millions of current dollars)*	*Employees (thousands)*
1961	2,010	63,802	145
1965	2,587	89,225	147
1973	5,723	177,878	209
1979	10,850	299,750	228
1987	63,771	2,284,166	516

Sources: Employees for 1961, 1965, 1973, and 1979 from NIPA, table 6.10B; for 1987, from *Survey of Current Business*, vol. 68 (July 1988), table 6.10B. Market values and number of shares traded for 1961, 1965, 1973, and 1979 taken from *Business Statistics, 1984, A Supplement to Survey of Current Business*, p. 75; for 1987, from *Survey of Current Business*, vol. 68 (July 1988).

Based upon number of shares traded per person engaged in the industry, productivity growth looks extremely strong in all years, accelerating in recent years. Based upon the value of shares traded deflated by the GNP deflator, productivity growth was weak over much of the period, but did accelerate post-1979. The figures are shown in table 17. These results certainly indicate that productivity growth has increased since 1979. Obviously many activities are being missed by this single output measure, including bond and option trading and the provision of investment advice. Taking account of these things would probably strengthen the conclusion that the industry has had rapid output per hour growth in the 1980s.

PRODUCTIVITY IN BANKING

Commercial banks engage in three main activities: transactions involving demand and time deposits, lending to businesses and consumers, including real estate loans, and fiduciary activities involving the management and administration of trusts and estates.[43] Other activities include money market operations for the bank's own portfolio, maintaining safe deposit boxes, issuing insurance, and giving investment

43. This section draws on Horst Brand and John Duke, "Productivity in Commercial Banking: Computers Spur the Advance," *Monthly Labor Review* (December 1982), pp. 19–27; Stephen Ledford, *BAI Survey of the Check Collection System* (Rolling Meadows, Illinois: Bank Administration Institute, 1987).

Table 17. Growth of Shares Traded per Employee in the Securities Industry, 1965–87, Selected Periods

Percent per year

Period	*Number of shares per employee*	*Real market value of shares per employee*[a]
1965–73	5.68	−0.54
1973–79	9.65	−0.46
1979–87	12.67	11.38

Sources: Same as table 16.
a. Market value adjusted by GNP deflator.

advice. By far the most important of these activities involves processing the transactions made with demand deposits and processing loans.

Substantial productivity increases in check processing have been revealed by surveys by the Bank Administration Institute from 1971 to 1986.[44] The Institute looks at several indicators of improved performance, and at overall labor productivity in particular. The number of checks processed per hour rose from 265 items per worker hour in 1971 to 825 items per worker hour in 1986, a rate of increase of 7.6 percent a year. During 1973–79 the rate of increase was 6.2 percent a year and during 1979–86 it was 6.3 percent a year. Clearly in check processing, a major activity of banks, there has been an enormous increase in output per hour. And this measure of productivity growth understates labor productivity improvement in handling transactions, because of the growth of electronic funds transfers (EFTs). In 1979 a conventional check cost 50 cents to process, whereas by contrast, EFTs had an average incremental processing cost of 7 cents.[45] At present EFTs represent only 1–2 percent of transactions; their quantitative impact is probably yet to be felt. But it is unlikely that productivity in the processing of demand deposit transactions will slow down.

THE BLS BANKING INDEX

The BLS now publishes an index of labor productivity for commercial banking based upon the check processing activities of banks and their loan and trust activities. Like all indexes in the BLS industry productivity measurement program, including those for transportation and retailing

44. Ledford, *BAI Survey of the Check Collection System*.

45. Frederick J. Schroeder, "Developments in Consumer Electronic Funds Transfers," *Federal Reserve Bulletin* (June 1983), p. 395.

examined below, the BLS banking index is quite separate from the BEA estimate of value added in the industry.[46] BLS combines the three types of activity into an overall output index using the shares of employment associated with each, based on periodic surveys by the Federal Reserve. The annual percentage growth rates in output per employee hour that they compute are as follows:[47]

Period	*Output growth*
1967–73	2.25
1973–79	0.45
1979–85	1.76

These numbers offer a striking parallel to productivity trends in the whole economy. Growth was fairly rapid before 1973, slowed almost to zero during 1973–79, and has made a partial recovery since then.

Why the sustained growth in the productivity of check processing did not boost productivity growth overall in the industry may seem puzzling, but there are three explanations. First, there has been a tremendous increase in branch banking. So many banks now dot street corners that Art Buchwald has suggested that they be required to sell gas to replace the gas stations they are displacing. Between 1967 and 1979 the number of bank offices rose 62 percent, and the average population served per bank fell from 6,000 in 1970 to 4,400 in 1980.[48] The link from the increase in branch banking to slower productivity growth is made on the basis of the "wide agreement among industry observers that scale economies in banking have declined with the spread of branching—that is, more resources including labor inputs, are required per unit of output."[49]

We are not persuaded that this "wide agreement" is quite as definitive as is suggested. The case for economies of scale is based on data showing the superior performance of large banks over small town and rural banks. But that superiority probably arose more from the relative organizational and technological backwardness of small banks. The trend in the industry has been toward consolidation of small banks into large companies,

46. The findings reported here are not incorporated at all in BLS's estimates of aggregate business productivity.

47. See Brand and Duke, "Productivity in Commercial Banking"; BLS, "Continued Gains in Industry Productivity in 1987 Reported by BLS," *News*, October 3, 1988.

48. American Bankers Association, *Statistical Information on the Financial Services Industry* (Washington, 1981), p. 89.

49. Brand and Duke, "Productivity in Commercial Banking," p. 25. They cite two "definitive" studies of economies of scale.

where the branches have access to the technology and computer facilities that were previously available only in large banks. Probably just as important is convenience, one of the main themes of our paper that emerges here and again in studying the retail industry. Clearly a reduction in the population-to-branch ratio must bring the average branch closer and reduce the shoe-leather transaction cost of consumer financial activities.

The second reason for the reported slow growth of banking productivity is that the BLS procedure of weighting individual activities by their relative labor input shares automatically biases against a finding of growth. Consider the case of consumer loans. Despite computerization, the processing of a conventional consumer loan takes a lot of time. In order to economize on the transactions costs involved in loans, banks now issue lines of credit as a matter of routine, including, of course, conventional credit cards. With a pre-approved line of credit, consumers can borrow and repay easily. And in response to changes in the tax law, this is now being done routinely for loans backed with real estate. Banks have been able to make loan transactions comparable in cost to that of processing checks. The BLS procedure makes these transactions into a separate category from conventional loans, and gives them a low weight precisely because of their low labor input.

The third reason for a slower than expected rate of growth in bank productivity is that computerization has been difficult, because of the history of the industry and its regulatory controls. The banking system has developed with small banks serving local markets. Each bank developed its own operating procedures, with the result that standard software packages cannot be simply applied to a given bank. The many special features written into the package create so many problems that bank staff often have to continue to use the old paper-based procedures in parallel with the new methods. Sometimes the dual procedures can compound the problems. For example, an account is closed by making computer entries and by having paperwork processed at the head office. By the time the paperwork is completed, the computer has charged a new service charge to the account and refuses to close it because the account has a negative balance.[50]

50. This discussion was based on a conversation with Frank W. Reilly, of Reilly Associates, Inc., a consultant to the banking industry. Software packages for Washington-area banks are written by local companies that have had difficulty remaining solvent because of repeated problems in getting bugs out of programs.

A ROUGH REESTIMATE OF PRODUCTIVITY IN FINANCE, INSURANCE, AND REAL ESTATE

The decline in the rate of return in real estate and the problems of computerization in banking help explain why productivity gains in finance, insurance, and real estate may have been restrained in the past 15 years. But every other indication suggests that productivity growth is being understated. The use of rental price indexes effectively carries some of the problems of measurement in the construction sector, to be discussed next, into this service industry. The deflation of the insurance industry is inappropriate and has biased the estimates of growth in all periods. The extent of the bias has probably increased because the procedures effectively deflate away all the gains that have come from computerization and the rapid spread of group insurance. The financial services industry has become extremely innovative in the past 15 years, and the BLS attempt to capture this in the banking industry is flawed in execution.

Although our analysis does not allow direct estimates of the errors involved in productivity measurement, it is worthwhile making an order-of-magnitude calculation of what might be involved. Banks, credit agencies, securities firms, and holding companies are assumed to have no growth. Suppose they had 2 percent growth before 1973 and 3 percent after 1973. The insurance industry has shown growth, but growth has been understated, especially since 1973. Suppose the understatement was 1 percent pre-1973 and 2 percent post-1973. The real estate sector showed over 4 percent growth before 1965, but nothing after. Problems with construction deflators surely contributed to this collapse. Suppose there were 2 percent a year understatement after 1965 (not just after 1973). Given the shares of the industries in the total sector, these magnitudes would imply a 1.1 percent a year understatement of growth pre-1973 and 2.3 percent understatement post-1973.

The Construction Enigma

There is no more obvious candidate than construction for a measurement explanation of the productivity growth slowdown. The basic

Table 18. Construction Industry, United States and Canada, 1948–86, Selected Years
Index, 1967 = 100

Item	*1948*	*1957*	*1967*	*1972*	*1977*	*1982*	*1986*
United States							
1. Structures real GNP	48.3	71.8	100.0	120.1	116.0	104.1	134.7
2. Real new construction	47.6	70.4	100.0	118.3	108.2	92.1	128.8
3. Construction materials	52.3	73.6	100.0	124.0	129.7	114.5	161.7
4. Real value added	47.2	74.7	100.0	87.4	86.6	73.9	88.3
5. Hours worked	68.0	84.1	100.0	111.0	112.7	113.1	140.3
6. Real net capital stock	52.8	69.1	100.0	124.7	157.2	164.4	n.a.
7. Output-hours ratio	70.0	83.8	100.0	106.6	96.0	81.4	91.8
8. Materials-hours ratio	76.9	87.6	100.0	111.7	115.1	101.2	115.2
9. Value added-hours ratio	69.4	88.9	100.0	78.7	76.9	65.3	62.9
10. Capital-hours ratio	77.7	82.2	100.0	112.3	139.5	145.3	n.a.
11. Materials-value added ratio	110.8	98.5	100.0	141.9	149.8	154.9	183.1
12. Construction deflator–GNP deflator ratio	106.5	108.9	100.0	105.7	113.2	104.5	114.9
Canada							
13. Construction output per full-time equivalent employee	62.0	87.4	100.0	97.0	105.4	126.0	121.3
14. Construction deflator–GNP deflator ratio	96.3	92.1	100.0	115.8	124.6	112.0	103.5

Sources: Line 1, NIPA, table 1.4. Line 2, NIPA, table 5.5, total less mining exploration, brokers' commissions, and mobile homes. Line 3, *Business Statistics* and *Survey of Current Business*, category of index of industrial supplies called "Construction Products" or, subsequently, "Construction Supplies." Line 4, NIPA, table 6.2. Line 5, NIPA, table 6.11. Line 6, John C. Musgrave, "Fixed Reproducible, Tangible Wealth in the United States: Revised Estimates," *Survey of Current Business*, vol. 66 (January 1986), pp. 51–76. Line 7, lines 2/5. Line 8, lines 3/5. Line 9, lines 4/5. Line 10, lines 6/5. Line 11, lines 3/4. Line 12 is line 1, table 19 below, divided by GNP deflator. Line 13, *Statistics Canada*. The figure listed under 1986 is, in fact, for 1985. Line 14, *Statistics Canada*.
n.a. Not available.

elements are shown in table 18, and the main fact to be explained glares out from line 9, taunting us with its implausibility. Construction value added per hour fell, so the data allege, by 20 percent between 1967 and 1972, and by another 20 percent between 1977 and 1986. The value-added and hours figures are shown separately on lines 4 and 5; between 1967 and 1986 real value added fell by 12 percent despite a 40 percent rise in hours worked. Since the weights of value added and of materials input in total construction output are roughly equal, that is, 50–50, the value-added and materials input index numbers on lines 3 and 4 roughly bracket the real new construction index number on line 2. This puts a helpful perspective on the productivity puzzle. "The reason" that real value added in 1986 is 73.4 index points below construction materials input is that the real new construction has an index number 32.9 points below construction materials input. Any resolution to the construction productivity puzzle, then, must involve boosting the 1986 index number on line 2 for real new construction, or reducing the 1986 index number on line 3 for construction materials input (a longstanding official component of the index of industrial production). Resolutions could be any combination of an undercount of nominal new construction, an upward bias in the construction deflator(s), and an overcount of materials.

Two factors might lead to an overcount of materials. First, the index of industrial production (IIP) is compiled from the suppliers' end and takes no account of the nature of the user. An increase in the share of materials going to users outside the market construction sector could cause the materials index in line 3 to overstate the increase in the use of materials by the market sector. We have already expressed skepticism that the underground economy has been growing, and we know of no evidence on the relative size of nonmarket home construction and handyman activity. A second factor is the spread of prefabrication, which is likely to cause the IIP to double-count the production of construction materials for final use. Consider a shift from doors built from raw lumber on-site to prefabricated doors built of exactly the same lumber. The economy's final use of doors is unchanged, yet the component of the IIP measuring the production of prefabricated doors will increase, while the production of lumber will stay the same. Correct practice would be to exclude production of lumber for intermediate use. We have no explicit evidence that prefabrication spread more rapidly after 1967 than before, but we suggest that the closely similar growth rates of materials and real construction output before 1967, in contrast to their sharply divergent paths after 1967, make prefabrication a plausible part of the needed explanation.

Turning now from materials to the measurement of construction output itself, we can put the task in an interesting way by asking, "How much would real new construction in 1986 have to be increased to eliminate the absolute 1967–86 decline in the level of productivity?" This would require an index number for value added equal to that for hours worked, 140.5, that in turn would require an index number for real new construction of about 151 instead of 128, or an extra 18 percent, taking materials input as given. Given the large rates of drift between alternative and official price indexes for producers' durable goods shown in table 9, the required bias for the aggregate construction deflator, 18 percent over 19 years or about 0.87 percentage points a year, seems

Table 19. Various Construction Prices, United States, 1929–86, Selected Years
Index, 1967 = 100

Item	*1929*	*1948*	*1957*	*1967*	*1972*	*1977*	*1982*	*1986*
1. Implicit structures deflator	35.6	70.0	88.3	100.0	136.9	212.1	332.9	365.3
2. Construction materials price index	n.a.	73.1	94.0	100.0	126.2	203.4	293.7	317.4
3. BLS average hourly earnings, construction	n.a.	41.7	65.9	100.0	147.4	197.1	283.0	303.6
4. Implicit deflator, residential structures	36.6	73.6	90.6	100.0	133.2	209.0	323.5	359.5
5. Price per square foot, residential	38.5	60.7	85.4	100.0	123.8	198.6	309.2	371.0
6. Implicit deflator, nonresidential structures	34.5	64.6	86.3	100.0	139.5	211.4	343.3	362.0
7. Törnqvist-weighted price per square foot, nonresidential	35.4	55.2	79.9	100.0	128.3	194.2	363.9	386.2
8. Implied quality per square foot								
a. Residential	105.2	82.5	94.3	100.0	92.9	94.9	95.4	103.2
b. Nonresidential	102.6	85.5	92.6	100.0	91.9	91.8	106.0	106.7

Sources: Lines 1, 4, and 6 from NIPA, table 5.5. Line 2 from *Economic Report of the President, February 1988.* Line 3 from BLS, *Employment, Hours, and Earnings, 1909–84,* vol. 2 (GPO, 1985), p. 912, and *Supplement to Employment and Earnings* (August 1988), p. 278. Line 5 from U.S. Bureau of the Census, *Statistical Abstract of the United States,* various issues, table titled "Construction Contracts." Line 7, authors' calculations. Line 8a is the ratio as a percent of line 5 to line 4; line 8b is the ratio as a percent of line 7 to line 6.

n.a. Not available.

small enough to be plausibly explained away in order to rid the economy of that perplexing productivity statistic on line 9. Any added contribution from an overstatement of materials input or an understatement of nominal construction output would make the task even easier.

Several of the remaining lines in table 18 indicate why the existing ratio of value added to hours is so implausible. First, workers are not less productive. They handled 15 percent more materials per hour in 1986 than in 1967 (line 8). Further, there is no reason why there should have been a substitution toward materials and away from labor. As shown in table 19, lines 2 and 3, labor actually became slightly cheaper over 1967–86 relative to materials. Second, the reduction in value added per worker does not represent a substitution away from capital toward labor. The capital-labor ratio increased substantially, at least through 1982. Finally, and perhaps most convincing, is the stark contrast between the Canadian productivity series on line 13 and the U.S. series on line 9. The nearly parallel movements of Canadian and U.S. productivity for 1948–67, followed by a sharp divergence, constitutes convincing circumstantial evidence of a measurement problem in the United States. The

Canadians have developed an extensive program to measure the prices of buildings by selecting prototype versions of five types of buildings. Rather than attempt to price the entire building, Statistics Canada prices the component parts of the building. Altogether about 100 different items are priced.[51] In principle this technique resembles the old U.S. "FHA-70 cities" index for residential housing that was compiled between 1947 and the late 1960s but then was discontinued.[52] One problem in reaching an easy conclusion that the whole explanation is inferior U.S. price deflators appears in table 18, lines 12 and 14, which show no marked divergence in the implicit price deflators for structures relative to the GNP deflator for the two countries. Clearly, further study of the Canadian data seems called for.

The task at hand is to eliminate the decline in construction productivity, which we have argued requires finding a bias in the overall structures deflator of roughly 0.9 percent a year for 1967–86. Several series related to price movements are assembled in table 19. Among the best hints is the behavior of the F. W. Dodge price per square foot series, shown separately for residential and nonresidential construction on lines 5 and 7. Because of a substantial shift in composition toward lower-value categories, line 7 is constructed as a Törnqvist index of seven separate categories of nonresidential construction, correcting completely for all available data on mix shifts. The ratio of the price per square foot series for residential construction to the existing implicit residential structures deflator, as shown on line 8a, constitutes an implicit index of quality per square foot. If quality increases, then the crude index of price per square foot should rise relative to an index that in principle corrects for all quality changes. The same quality proxy is shown for nonresidential construction in line 8b.

51. The source for this description of the Canadian methodology is Paul Pieper, "The Measurement of Structures Prices: Retrospect and Prospect," paper presented at 50th Anniversary Conference, Conference on Research in Income and Wealth, Washington, D.C., May 12–14, 1988.

52. The annual growth rate of the FHA index over 1947–68 was 1.68 percent. This does not provide any indication of the possible bias of the bid-price technique compared to alternatives, since the FHA index is the basis for half of the NIPA residential structures deflator over the period 1947–63. The annual growth rate of the implicit NIPA structures deflator over 1947–68 is 2.37 percent. Comparisons of the FHA index with various backward extrapolations of the Census hedonic price index for single-family houses, used in the NIPA since 1963, are in Robert J. Gordon, "An Evaluation of Alternative Approaches to Construction Price Deflation," unpublished consultant report, August 1969.

Two aspects of the implicit quality indexes strike us as remarkable. First, given the very different procedures by which the residential and nonresidential deflators are compiled, the similarity of the implicit quality indexes is surprising. But more important, and in our view strong circumstantial evidence against the deflators, is the conclusion that there was absolutely no improvement in the quality per square foot of either residential or nonresidential structures from 1929 to 1986. Two approaches seem open to deal with the possibility of deflator bias. One, recently carried out by Paul Pieper, is to study the components of the official deflators directly and examine their sensitivity to alternative methodology. The second is to compile direct evidence on the quality of both residential and nonresidential structures.

Almost all the quantitative evidence available on the quality of structures refers to single-family residential housing. Since 1963 the NIPA residential structures deflator has been based on the Census Bureau hedonic price index for single-family homes (the only other hedonic price index in the NIPA besides the computer price index); any claim that quality improvements are understated must thus be based on some flaw in this index. The hedonic regression equation explains most of the cross-sectional variance of house prices by a single square-foot variable and implicitly includes in this coefficient the value of those features that typically vary between large and small houses, including number of bathrooms, size of furnace, and so forth. However, such an index can be biased if features are now included in relatively large houses that were not included two decades ago in houses of the same size. For instance, starting from zero in 1948 and 29 percent in 1967, the share of new houses built with central air conditioning reached 72 percent in 1981.[53] The share of built-in dishwashers grew from 45 percent to 82 percent, with a roughly similar performance for garbage disposals and ventilator fans. Real production of wood kitchen cabinets increased 157 percent between 1967 and 1980, compared with a 50 percent increase in real expenditures on new residential housing and on additions and alterations. Data indicate a marked increase after 1974 in double-glazed windows and incidence of wall insulation. In contrast to 1969, when a

53. Data in this section come from Paul Pieper, "The Measurement of Real Investment in Structures and the Construction Productivity Decline" (Ph.D. dissertation, Northwestern University, June 1984).

BLS survey indicated that only 0.2 percent of the construction cost of the average home was devoted to nursery products and landscaping, Chicago suburban developers recently were quoted as spending 2–4 percent on landscaping.[54] The two main dimensions of quality deterioration in residential housing, the shift from plaster to drywall and from hardwood floors to wall-to-wall carpeting, were largely complete by 1967.

Pieper compiles construction indexes for both residential and nonresidential structures by several alternative methods. His own preferred index yields a bias or drift for total construction of 0.6 annual percentage point for 1963–82, 0.3 point for 1963–72, and 0.8 point for 1972–82. Most of the effect comes from his residential index, which simply involves recalculating the Census hedonic price index for single-family houses to exclude the largest houses, of a size 2,400 square feet and over.[55] The argument is that the largest houses are most likely to be custom built and to have an increasing ratio of amenities to square feet of area, which seems to be the main quality characteristic captured by the Census hedonic regression. It is quite likely that Pieper's adjustments are too small, since in previous work he has pointed to numerous additional sources of bias in the Census price index that are not taken into account in the more recent study. Among the most important of these are a downward bias in the Census estimate of the portion of the house price constituting land value, which must be excluded from the observed house price to determine the price of the structure itself. Another problem is a shift in the regional composition of construction toward lower-cost areas, which is implicitly treated by the Census technique as a quality decline.

Overall, it seems highly likely that the residential deflator is upward biased by at least 1.0 percent a year since 1963, given the number of factors that Pieper's 0.6 percentage point bias estimate fails to take into account. Since the 1967–86 increase in the implicit nonresidential structures deflator (table 19, line 6) is almost identical to that of the residential structures deflator, it is quite likely that the former is biased at least as much. While we have not come across any explicit evidence on quality of nonresidential buildings, casual observation suggests a significant

54. "Green Power: Landscaping's Appeal Grows," *Chicago Tribune*, July 30, 1988.

55. Pieper notes that the price increase during 1974–81 for large houses in the Census sample was almost double that for houses with 1,000–1,200 square feet.

improvement in the quality and type of materials included in the typical new retail store, hotel, or even university classroom compared with its counterpart of 20 or 40 years ago. The typical new office building is constructed with intermediate layers between stories, allowing a high degree of flexibility and access for electric lines and cooling ducts. Heating and air conditioning control systems are more sophisticated, and elevators are faster. The implied verdict of the existing deflators that the quality of nonresidential buildings has increased by just 7 percent since 1967 seems dubious at best.

As has been true for many years, it is hard to convert this set of circumstantial evidence into a case that would stand up in court. As the next step in the research, a close look at the underlying ingredients of the Canadian deflators and their implications for quality change, particularly of nonresidential buildings, could have a high payoff.

Retail Trade

The provision of convenience can be interpreted as a technological change that allows the substitution of low-value for high-value hours of the day or week. Some evolutions in the service industries may represent the production by computers of convenience (24-hour money machines and bill-paying services) that does not enter GNP; others may represent the production of unmeasured convenience in ways that reduce measured GNP per employee (24-hour convenience stores). In discussing convenience, we begin by distinguishing between three uses of time: work that pays the market wage, home time that yields direct utility, and transaction time that is an evil needed to produce both work time and home time. The value of a technologically induced reduction in time required to carry out transactions is measured by the market wage. If laser scanners in supermarkets reduce average waiting time five minutes, the value of the five minutes is equal to the average wage that could be earned on an extra five minutes of work. If laser scanners allow both a reduction in supermarket employees and a reduction in waiting time, then the value of both should be included in measures of output per supermarket employee.

The value of home time can be differentiated by time-of-day and time-of-week. A family that places a high value on eating dinner together or watching the 11 p.m. news together may value those particular hours of home time highly, while valuing less highly 2 p.m. on Sunday, when they have "nothing to do." The value of 24-hour money machines, convenience stores, liberal store closing regulations, and VCRs can be interpreted as allowing the optimal shifting of transactions activities to lower-value hours. To provide an extreme example of the value of convenience: rigid shop-closing hours in Germany cause a given level of goods production to be sold by fewer people in a short period of time, leading to a spuriously high level of market productivity in the German retailing sector, while no account is taken of the congestion, extra transactions time spent waiting in line, and high-value home time unnecessarily spent in transactions. German regulations require virtually all retail establishments to close at 6:30 p.m. on weekdays and 1:30 p.m. on Saturday afternoon, with no Sunday opening at all. If such regulations were applied in the United States and they caused five hours a week to be shifted from $3 an hour time to $8 an hour time, the value over the entire U.S. population (aged 16 and over) would be $239 billion, or about 5 percent of 1987 GNP. To place this number in perspective, value added by the U.S. retail sector in 1982 was about 9 percent of GNP, so clearly even a modest allowance for convenience-shift effects might offset much if not all of the productivity slowdown in retailing. The problem, as we shall see, is that the spread of convenience long antedates the productivity slowdown, and in the case of supermarkets actually works against an explanation of the productivity slowdown.

RETAIL TRADE PRODUCTIVITY BY TYPE OF STORE

Retail trade makes a relatively small contribution to the productivity growth slowdown in the BLS data, as is evident in table 4 above. Since retail trade is the most extensively covered area of the services in the BLS industry productivity measurement program, it is interesting to compare the BEA and BLS estimates of productivity growth and to see in the more detailed BLS data which types of retailing are doing well and which badly. Once again, we stress, as in the discussion of banking above, that no use is made of the BLS industry data, either by the BEA in measuring industry output or by the BLS itself in its measures of aggregate productivity. In fact, as far as we can determine, the BLS

Table 20. Retail Trade, Productivity Indexes for the Aggregate and for Eight Types of Stores, 1958–86, Selected Years

Index, 1967 = 100

Item	*1958*	*1967*	*1972*	*1977*	*1982*	*1986*
1. BEA output per hour, retail trade	87.3	100.0	106.0	109.3	112.0	121.9
2. BLS output per hour, retail trade	78.2	100.0	115.3	121.4	125.9	139.1
3. Food stores	75.3	100.0	112.6	102.0	95.4	94.9
4. Eating and drinking places	91.8	100.0	104.9	102.6	99.1	98.8
5. Total food	80.2	100.0	110.3	102.2	96.5	96.0
6. Department stores	n.a.	100.0	107.8	128.4	137.9	167.3
7. Apparel and accessories	n.a.	100.0	111.4	121.8	159.4	198.3
8. New car dealers	76.9	100.0	117.2	123.5	124.0	137.4
9. Gas stations	77.3	100.0	128.0	158.2	175.2	200.3
10. Furniture, home furnishings, and equipment	n.a.	100.0	126.9	142.4	155.6	201.2
11. Drug and proprietary stores	68.1	100.0	131.8	149.3	161.0	144.8
12. Total nonfood	76.9	100.0	118.6	133.9	145.1	167.2

Sources: BEA: NIPA, table 6.2, divided by table 6.11B. BLS: *Productivity Measures for Selected Industries and Government Services*, BLS Bulletin 2296, February 1988, pp. 143–50. For share weights of components used in computing totals in lines 2, 5, and 12, see *Business Statistics, 1984*, p. 37.
n.a. Not available.

never aggregates its industry productivity measures as we do here and in the final case study.

Table 20 begins with BEA output per hour in retail trade and provides a contrast with the BLS measures, both aggregate and for eight types of retail stores, separated into food and nonfood. The BLS index for food store productivity is shown on line 5, for nonfood on line 12; total retail productivity appears on line 2. The BLS indexes cover store-types accounting for fully 82.6 percent of total 1982 retail sales. What pops out immediately from the page is the stark contrast between the experience of food and nonfood retailing since 1972. There has been a sufficient productivity growth slowdown in food retailing to qualify this sector as a "basket-case" industry, along with construction, mining, and electric utilities. For nonfood retailing there has been virtually no slowdown. The annual rates of growth corresponding to the levels in table 20 are shown in table 21. The failure of the BLS index on line 4 of table 20 to record any increase in the productivity in eating and drinking places would make Ray Kroc, the founder of McDonald's, turn in his grave. Even without explicitly treating the extended hours of service and other

Table 21. Growth of Real Output per Hour in Retail Trade, 1958–86, Selected Periods
Percent per year

Retail index	*1958–72*	*1972–86*	*Change, 1958–72 to 1972–86*
BEA total	1.39	1.00	−0.39
BLS total	2.77	1.34	−1.43
BLS food	2.28	−0.99	−3.27
BLS nonfood	3.09	2.45	−0.64

Source: Calculations based on table 20.

aspects of consumer convenience made possible by franchised fast food outlets, we suspect that there is room for significant errors in deflating consumer expenditures on food. Relatively small quality differences seem to justify substantial relative price differences in this industry, and it is quite likely that the CPI may miss quality changes over time. For instance, the typical fast food hamburger is not the same commodity as it was 20 years ago, because more and more fast-food restaurants have become sit-down restaurants, with salad bars no less, and have subtly shifted into a higher price category.[56] Pending detailed discussions with BLS experts, we leave these thoughts as conjectures.

We are on firmer ground for stores selling food for consumption at home, that is, supermarkets, because substantial quantitative data are available on changes in the nature of supermarkets over time. The puzzling aspect of the absolute decline in food store productivity is its dramatic contrast with the three types of nonfood stores where productivity has doubled since 1967—gas stations, furniture stores, and apparel stores. Special factors may explain the first two. Gas stations have been affected by the shift to self-service, which arguably represents an unmeasured decline in quality, and by a marked increase in the population of automobiles per available gasoline pump.[57] Furniture stores, which

56. We would be extremely surprised if exactly the same point were not even more true for hotels. In 1987 a survey of 659 hotels showed the following percentages offered these guest services: free parking (81 percent), personal bathroom amenities (80 percent), audio-video equipment (75 percent), outdoor pool (61 percent), free airport transportation (52 percent), and health club facilities (45 percent); see "Hotels Fight Vacancies with More Services," *Wall Street Journal*, October 3, 1988.

57. The 1982 *Census of Retailing* reports that the number of gas pumps in the United States declined 28 percent between 1972 and 1982, and that the fraction of self-service pumps (which was not reported in 1972) was 54 percent in 1982.

Table 22. Indicators of Productivity and Services Provided, U.S. Supermarkets, 1948–85, Selected Years

Index, 1967 = 100, except as noted

Item	*1948*	*1957*	*1967*	*1972*	*1977*	*1982*	*1985*
1. Real sales per hour	58.2	82.0	100.0	106.4	100.8	103.6	92.7
2. Real sales per square foot	n.a.	121.6	100.0	97.5	83.6	77.5	75.6
3. Real sales per transaction	n.a.	86.0	100.0	98.4	94.0	93.3	97.0
4. Square feet per hour worked	n.a.	67.4	100.0	109.1	120.6	133.7	122.6
5. Square feet per member of U.S. population, 16 and over	n.a.	77.4	100.0	110.3	117.1	123.3	128.8
6. Items carried per store	2,200	4,800	7,000	9,000	10,500	13,067	17,459
7. Percent stores with complete air conditioning	14	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
8. Percent with mechanical refrigeration in produce department	30	n.a.	n.a.	n.a.	100	n.a.	n.a.
9. Percent with "extensive" delis	n.a.	n.a.	46	n.a.	n.a.	n.a.	68
10. Percent selling beer	29	47	55	n.a.	n.a.	n.a.	80
11. Percent open Sunday	5	24	45	55	62	n.a.	n.a.
12. Percent open seven evenings	22	51	73	n.a.	n.a.	n.a.	n.a.

Sources: Supermarket data from annual issues of Super Market Institute, *The Super Market Industry Speaks* (Chicago, 1986). Sales deflated by CPI for food at home. 1948 figure for items carried refers to 1950.
n.a. Not available.

sell electronic goods and appliances, have benefited from a marked increase in real value added per transaction. Apparel stores remain as the remarkable success story that raises questions about the food sector, for the two share many physical operations, including product storage, shelving, and check out, and bear the same stamp of computers, in the form of scanners and point-of-sale terminals.

What is the problem in the supermarket industry? The top section of table 22 reports several productivity measures, all of which show declines since 1967. Real sales per hour closely mimic the BLS index for food stores in table 20, increasing at a healthy pace through 1972 and then declining, while real sales per square foot decline at an erratic pace throughout 1958–85. Real sales per transaction are essentially trendless. The data shown imply a substantial increase in square feet per hour, that is, an increase in the capital-labor ratio, at least until 1982, and a roughly similar growth in supermarket square feet per member of the adult population.

The purpose of all these square feet becomes apparent in line 6, which shows an explosion in items carried per store from 2,200 in 1950 to 17,500 in 1985. This growth in the variety of goods available provides convenience to the consumer by reducing total transactions time. Some of the scattered figures presented in the bottom of the table document the increasing variety of goods available in the typical supermarket that cut the number of shopping trips required by the average consumer to purchase a given market basket, for example, deli and fresh fish items. Also shown are direct measures of the quality of the supermarket itself and of the extension of opening hours, which directly reduces productivity by spreading food shopping over a greater amount of labor input.

These factors, however, do not appear to contribute an explanation of the productivity slowdown. Every factor, including the increase in variety of items and the extension of hours, was more rapid in the first half of the postwar period. It appears from these numbers that most of the productivity growth slowdown in the supermarket industry is real, an example of Nordhaus's depletion hypothesis.[58] We could invent the supermarket only once, just as we could build the interstate highway system only once. Still, something else has been happening that represents at least a modest role for mismeasurement, and we believe that it can be documented in due course. After moving almost completely away from service to self-service in the first half of the postwar period, supermarkets are moving back to service. There is scattered evidence of a movement back toward labor-intensive services valued by consumers, including full-service deli and seafood counters. Most supermarkets that have salad bars allocate at least one full-time individual to take care of them. Interestingly, most supermarkets have a full-time individual in charge of the electronic scanners, the main purpose of which seems to be information gathering rather than improving productivity.[59]

Transportation

Our final case study illustrates the last theme, the need for consistency across government agencies and the role of inconsistent data as a flag to

58. William D. Nordhaus, "Economic Policy in the Face of Declining Productivity Growth," *European Economic Review*, vol. 18 (May–June 1982), pp. 131–58.

59. Over half of supermarkets use scanner data for checker performance, tracking the results of special promotions, coupon accounting, and work scheduling. Other uses are meat and produce department analysis, new product evaluation, shelf allocation, "shrink analysis," and, mentioned by 21 percent of respondents, "price elasticity." This list and the percentage responses are from Super Market Institute, *Detailed Tabulations* (Chicago, 1986), table 78B.

Table 23. Real Growth of Output per Employee, Transportation Sector, 1948–86, Selected Periods[a]

Percent per year

Index	*1948–72*	*1972–86*	*Change, 1948–72 to 1972–86*
BEA	3.56	0.38	−3.18
BLS	3.74[b]	2.01	−1.73
Physical traffic measures	3.32	3.29	−0.03

Source: Törnqvist averages from table 24 below.
a. Measured by change in Törnqvist average.
b. BLS index not available for 1948–58. Extrapolated backwards using BEA index.

indicate the need for improvements in measurement. Since this case study is mainly in the nature of a report rather than an analysis, we present the overall results first and then discuss them. To assess the BEA industry measures of transportation output per employee, we have developed a comparison with the BLS industry measures for the three main modes (rail, truck, and air, accounting for over 75 percent of transportation value added in 1982). Since there is a sharp disagreement between the BEA and BLS measures for rail and air, we have also compiled a simple index of traffic (ton-miles for rail and truck, passenger-miles for air) for comparison. As seen in table 23, the measured growth rates of output per employee show a sharp productivity growth slowdown for BEA, a substantial slowdown but four times faster productivity growth after 1972 for BLS, and no slowdown at all for the physical traffic series. The productivity series for the three modes are aggregated using the Törnqvist method, with nominal value-added shares in each subinterval as weights. Since the BEA and physical measures are available back to 1948, but the BLS measure only to 1958, we push the latter back by linking it to the BEA measure. While this makes the close accordance of the 1948–72 BEA and BLS growth rates partly spurious, the actual measures do grow at absolutely identical rates between 1958 and 1967, before diverging after 1967.

Turning to the detailed productivity growth rates in table 24, we can contrast the modest BEA 2.3 percent annual growth rate for railroads over 1972–86 with the much faster BLS rate of 4.7 percent and the physical index rate of 5.4 percent. For airlines the BEA figures are in even greater disagreement, registering a *decline* at an annual rate of 0.2 percent compared with increases for the BLS of 3.6 percent and the

Table 24. Alternative Productivity Growth Rates for Selected Transportation Industries, Output per Employee, 1948–86, Selected Periods

Percent per year

Item	*1948–58*	*1958–67*	*1967–72*	*1972–77*	*1977–82*	*1982–86*
BEA						
1. Railroads	2.16	6.44	1.35	2.24	−2.32	8.15
2. Trucking	2.76	3.06	4.32	2.20	−2.55	0.06
3. Air transportation	7.11	5.08	3.02	1.89	−4.14	1.97
4. Törnqvist average	2.70	4.75	3.15	2.14	−2.85	2.20
BLS						
5. Railroad revenue traffic	n.a.	6.11	3.79	2.34	1.92	11.07
6. Intercity trucking	n.a.	2.42	3.80	1.13	−1.39	1.39
7. Air transportation	n.a.	8.04	5.05	3.54	2.77	4.61
8. Törnqvist average	n.a.	4.75	4.04	1.97	0.32	4.18
Physical output per employee						
9. Railroad freight ton-miles	2.80	5.77	4.49	2.95	4.14	10.03
10. Intercity trucking ton-miles	2.76[a]	0.82	1.03	1.32	1.65	4.09
11. Scheduled airline passenger miles	8.82	5.23	6.94	4.28	4.51	2.51
12. Törnqvist average	3.20	3.50	3.26	2.39	2.88	4.92
Addendum: Airline price index, final year of each period, 1967 = 100						
13. BEA PCE deflator for air transportation	72.8	100.0	131.2	193.2	400.0	381.2
14. Fixed-weight index, revenue passenger coach-class yield, domestic and international U.S. scheduled industry	112.8	100.0	127.0	153.4	225.3	205.3

Sources: BEA data from NIPA, tables 6.2 and 6.10. BLS data from BLS Bulletin 2296, pp. 134–38. Physical output: railroads from *Railroad Facts,* various editions; trucking industry employment and intercity freight by mode from *Statistical Abstract,* table titled "Class I Intercity Motor Carriers of Property, by Carrier," various issues; airline output and employment from Air Transport Association, *Air Transport,* various issues.

n.a. Not available.

a. Trucking not available for 1948; extrapolated backwards by BEA index from line 2.

physical measure of 3.9 percent. For trucks the three respective rates are much lower, and the BEA and BLS are closer together: −0.1 percent and 0.3 percent, respectively, as contrasted to 2.2 percent for the physical output measure.

While we have not determined the cause of the railroad and trucking discrepancies, the cause of the BEA's error in measuring airline output is amazingly simple—the BEA (or perhaps the CPI division of the BLS) forgot to take account of airline discount fares. This fact is documented at the bottom of table 24, where we compare the implicit NIPA deflator for consumer purchases of air transportation with a fixed-weighted index of domestic and international coach air "passenger yields," that is, revenue collected per passenger-mile. The BEA deflator implies that air fares almost tripled from 1972 to 1986, while the yield index suggests that the actual price paid increased by about 60 percent.[60]

The BLS indexes may be preferable to the physical output measures, because they take into explicit account shifts in mix between high-value and low-value freight. But in the context of our earlier discussion of computer services, the simple physical measures raise as an additional measurement issue the net effect of changes in the quality of a ton-mile or passenger-mile. Railroad and trucking services are mainly intermediate goods, so that this question is moot for the aggregate productivity slowdown, except insofar as computers improve the working lives of railroad and trucking employees. But airline services are purchased by consumers as well as businessmen. Most of the perceived quality improvement in airline services has been contributed by the computer, as discussed above, and much of any perceived deterioration due to air traffic delays can be attributed to the lack of parallel investment in new air traffic control computers by the Federal Aviation Administration. Because the BLS and physical output measures for airlines agree so closely, any remaining scope for mismeasurement of the quality of a passenger-mile will need to collect detailed evidence and weigh factors on both sides. Our final proviso is that any such study should include output and labor input in travel agencies, since these have taken over much of the former role of airline reservations agents, leading to at least some upward bias in all the airline productivity measures examined here.

Conclusion

We have uncovered large potential errors in the measurement of productivity at the industry level. However, as we have stressed since the beginning of the paper, the discovery of measurement errors at the industry level merely reshuffles the locus of productivity growth unless it can be shown that data on aggregate real output and (for ALP) labor input are subject to error. Further, some of the errors we have discovered

60. We are aware that passenger yield varies by length of haul and would be biased downward if there were a substantial increase in length of haul. However, between October 1977 and October 1987, length of haul actually decreased slightly for the U.S. domestic trunk airlines, from 833 to 816 miles. Source: monthly traffic reports as published in *Aviation Week,* various issues.

may have applied not just in the slowdown period, but in the pre-slowdown period. This is particularly true of our study of construction prices, since the available data on quality per square foot provide evidence of improvements in quality per square foot before the advent of the productivity slowdown, not just during the slowdown period.

In this concluding section we limit our overall assessment to measures of ALP, since MFP measures require a more complex computation of effects on both output and capital growth, with considerable sensitivity to the weights chosen for capital's income share. At the level of ALP, we can classify our discussion of measurement issues by quadrant on the grid introduced at the beginning of the paper. We are looking for errors in the critical northwest quadrant, errors that matter in the aggregate economy and contribute to the slowdown. We have demonstrated effects on aggregate output of deflation errors in a wide range of producer and consumer durable goods, but these do not help, because they are in the northeast quadrant, indicating roughly the same degree of GNP mismeasurement before as after 1973 (in table 10 the top section for GNP indicates almost identical errors before and after 1973; there is a small contribution in the right direction for MFP shown at the bottom of that table). Pointing in the right direction are the effects of computers, which we have documented for several components of finance, insurance, and real estate, and the deflation error for consumer purchases of airline services.

Taken together, however, our specific quantitative contributions in the northwest corner are relatively small. The combined share of consumer expenditures on business services and on airline transportation is 4.1 percent of 1982 GNP, so even an outside estimate of a positive 3 percent annual contribution to the slowdown would add only 0.12 percentage point to productivity growth. We conjecture that further study could raise this to 0.2. The labor quality literature adds a little more. Denison found a contribution of labor quality to understanding the slowdown of about 0.1 point, which with a dose of Jorgenson and his colleagues might be raised to, say, 0.2 point, and Bishop's test data, we suggested, could add another 0.1 point to that. On balance, then, measurement issues, at least those that we have quantified to date, could account for only about 0.5 point of the full 1.5 point slowdown, mostly from the labor input side.

Other important parts of GNP that could be sources of measurement error are not necessarily located in the northwest quadrant of our grid. Clearly any payoff must come from finding measurement problems in nondurable consumption (24.3 percent of GNP in 1982) or services (32.4 percent). But measurement errors could just as well belong to the northeast corner; for example, errors in measuring the benefits of synthetic fabrics or in the improved quality of medical care could apply just as well before 1973 as after.[61] The same is true for our case study of construction; we concluded that the absolute decline of construction productivity after 1967 was likely to be spurious, but there is no good reason to think that most of the same biases did not occur as well in earlier years, leaving the productivity growth slowdown in construction intact, even if at a higher level.

The payoff to studies of measurement issues appears to be much greater at the industry level. Our new price deflators for durable goods radically change all the standard data on the ratios of investment and capital to GNP, the relative prices of investment and consumption goods, and the relative productivity of durable manufacturing to other parts of the economy. We find that non-goods-producing industries like communications, utilities, and transportation have been credited for MFP gains more appropriately counted as the payoff to research and innovation in durable manufacturing. All studies of investment and the payoff to R&D, among others, are affected by these results and will need to take them into account.

An interesting aspect of productivity in the postwar period emerges from our case studies. Between 1948 and 1967 most industries exhibit highly similar productivity growth rates; after 1967 this unified advance fragments. The cross-industry *variance* of productivity growth is much greater in the slowdown period, even at the disaggregated level. Railroads and airlines boom, while trucking slumps. Apparel and furniture stores become more productive rapidly, while productivity falls at supermarkets. Construction productivity goes down, while construction materials input goes up. This pattern seems consistent with the view of a common

61. Or a careful study of clothing quality might point in the opposite direction. See Francine Schwadel, "Complaints Rise about Clothing Quality," *Wall Street Journal*, June 27, 1988.

impetus to productivity advance in the early postwar years, perhaps a backlog of innovations and investment opportunities delayed by depression and war, followed, after the mid-1960s, by a depletion of opportunities and a reversion to more normal differentiation of each industry segment.

Finally, we have developed substantial quantitative evidence showing that ALP growth in particular sectors and industries—chief among them banking, securities, insurance, railroads, and airlines—has been substantially understated in the BEA data, particularly after 1973. In fact, the fragmentary data we have examined for banking and securities suggest that these industries may well have experienced some of the fastest rates of growth of ALP in the entire economy, next to computers and other electronic durable goods. As we carry out these crude pilot studies, we are struck with the potential for much fruitful work at the detailed industry level, and we must express some frustration in the present lack of coordination between the BEA and BLS in this regard. The BEA should examine the consistency or lack thereof between its industry output indexes and those developed in the BLS industry studies, as in transportation. The BEA should take into account the research effort that the BLS has devoted to such industries as banking. The shoe is on the other foot when it comes to price measurement. Why should the PPI fail to incorporate the BEA's computer price index? And how long must users of government statistics put up with the total lack of any PPI for the single most important component of PDE, communications equipment, when the PPI contains literally hundreds of detailed commodity indexes for nuts, bolts, pipes, flanges, valves, cans, barrels, pails, tanks, hinges, cleats, knives, and other crude products of lesser economic importance?

Comments and Discussion

William D. Nordhaus: Martin Baily and Robert Gordon have written a highly informative paper on the productivity growth slowdown, emphasizing the role of potential measurement errors, especially those relating to the growing use of computers. In the end, they find that measurement errors explain little of the slowdown, but along the way they uncover many fascinating and controversial problems.

At the outset, I would like to register a concern about the surprisingly ahistorical stance of economists toward the slowdown in productivity growth, which we are analyzing as if it were a slowdown in the speed of light. But surely there is nothing automatic about 3.2 percent per year growth in labor productivity. A glance backward over this century shows that U.S. productivity growth slowed significantly at least twice: productivity growth was absent from 1901 to 1917 and was extremely modest from 1924 to 1937.[1] Moreover, other countries have shown widely varying productivity experience and have also experienced sharp changes in productivity growth from decade to decade.

To put this point in a different way, given that Solow and Denison were unable to identify the sources of productivity growth in the first place, we should not be surprised if the growth disappears. Productivity growth is no more mysterious at 1 percent per year than at 3 percent.

Turning to the results of the paper, I find myself largely in agreement on most points. In my remarks, I will focus on some of the areas where their analysis seems a bit off target. One of the major puzzles of recent years is why the tremendous increase in computer power and use has not caused more productivity growth. Because computers are largely an

1. See William D. Nordhaus, "Economic Policy in the Face of Declining Productivity Growth," *European Economic Review*, vol. 18 (May–June 1982), pp. 131–58.

intermediate product, the mismeasurement issue in computers involves understatement of both output and input.

On the output side, although much has been made of the phenomenal growth in manufacturing productivity during the 1980s, the role of computers in that surge has been largely ignored. In fact, the enormous growth in computer output has been largely responsible for the rebound in manufacturing productivity over the past decade. The following, taken from the authors' table 4, shows the average annual percentage growth in labor productivity in three subperiods:

Sector	*1948–73*	*1973–79*	*1979–87*
Manufacturing	2.87	1.43	3.49
Manufacturing less nonelectrical machinery	2.96	1.53	2.19

Fully two-thirds of the rebound in manufacturing productivity growth after 1979 was due to the rapid productivity growth in computers (which is part of nonelectrical machinery).

A second feature of inclusion of computers is the potential error in aggregation of capital. Measured by the revised hedonic index, the share of computers in constant-dollar investment rose from 1.6 percent in 1969 to 25.3 percent in 1986. We should be cautious about productivity measures that simply aggregate computers and other capital goods into the stock of K to be used in productivity measures. A simple aggregation would violate the capital-aggregation requirement that relative capital goods rentals are unchanged. Using conventional data on the share of computers in the capital stock, a simple calculation indicates that 10 years after the base year, the growth of capital could be overstated by a factor of two if the standard technique for constructing total factor productivity is used. The underestimate of total factor productivity growth might be as much as half a percent a year.

Paradoxically, if we are to use an overall capital aggregate rather than appropriately weighted subaggregates, the old "naive" measurement of computer inputs (which approximately measures computer prices by the prices of other capital goods) would provide more accurate estimates of computers' contribution to productivity growth.

One of the themes that runs through Baily and Gordon's study is that there is a large component of unmeasured quality change in services. Two particular aspects are stressed: the change in the number of commodities and changes in the amenities associated with services.

While I tend to agree that we have understated the quality change associated with products, Baily and Gordon may overstate the quality change in product variety and services. Start with the first example of quality change, the increase in the number of commodities. They cite the example of grocery stores, where the number of goods has risen from 2,200 in 1948 to 17,459 in 1985. To assess the expanding choice, I went to the local grocery store. I found that about 5 percent of the shelves were devoted to cereals, including dozens of brands such as Freakies, Mueslix, Twinkies, Kix, Kasli, Total, Life, and Just Right. You might ask yourself how much your net economic welfare would decrease if the number of cereals were reduced by half.

The other example of quality change used by the authors is the amenities associated with services. Here, I would think that the contrary case is pretty strong. I am reminded of the visiting European who recently left his shoes in the hallway of the Holiday Inn to get shined. He was awakened at two o'clock in the morning by the house detective who advised him to retrieve his shoes or risk going barefoot in the morning.

Just to recall some of the deterioration in the quality of services, you may want to remember the doctor's house call, the butcher's custom steaks, and the tailor's custom suits. Yale men had daily maid service and sit-down meals, and undergraduates were even taught by the faculty rather than by graduate students. If you bought gasoline, you could expect to come away with a clean windshield and free maps rather than dirty hands. And who has recently enjoyed a gourmet meal on Amtrak? In short, the authors have a strong case when they argue that many *products* have undergone unmeasured quality change; but to extrapolate that argument to *services* requires excessive suspension of disbelief.

But let us assume they are right about quality change. Are we better off because of all the proliferation of Corn Pops and Freakies? The issue of the optimal amount of product differentiation is a profound one, and industrial organization economists reason that even if tastes are not manipulated, a market economy can easily produce excessive quality change because of the setup costs of product differentiation. The appropriate test here is to ask what tax or subsidy you would require, over the existing official price increase, to shop in a modern grocery store, to ride on Amtrak, to get your hair cut, as compared with the same service a few years ago?

Finally, say that you believe that the quality of services of the butcher, the baker, and the candlestick maker has improved. To contribute to an explanation of the productivity slowdown, there must be an *increase* in the growth of unmeasured service quality change since 1973.

In the area of construction, the puzzle of falling productivity has long plagued productivity experts. I find myself unswayed by the paper's reasoning, which basically dismisses the result as "taunting us with its implausibility." There is a circular reasoning that dismisses the results because of incredibility, while finding nothing other than some mysterious Canadian data to substitute for existing approaches.

At the end of the day, the authors conclude that the quantifiable measurement errors explain only 0.5 percentage point of the 1.5 point slowdown, with 0.3 point coming from deterioration in the quality of labor, 0.1 point coming from the mismeasurement of airline productivity, and another 0.1 point coming from a "conjecture" that "further study" will yield another 0.1 percentage point.

The 0.3 point from labor is quite fragile. First, two of the three studies show contrary results, and only the Jorgenson study shows a significant decline in labor quality. Second, as the paper explains, the Jorgenson study stops before the apparent upturn in the returns to education occurred, and that study attributes lower productivity to shifts in occupation and industry. Third, the estimated impact of lower test scores raises questions because the increase in test scores may reflect the increasing obsession with studying for test-taking techniques rather than studying for content; to the extent that it reflects a genuine deterioration in ability, that decline should already be captured in the age variables, with the lower productivities of later cohorts reflected in lower relative wages; and the increase in the coefficient is an indefensible statistical procedure.

Of the 0.2 percentage point from airlines and other consumer expenditures, it should be noted that the increase in airline productivity comes during 1979–87, so it can hardly explain the early part of the slowdown. And the conjectures about further study cannot substitute for evidence. Pending further evidence, I believe that the only reliable evidence of a mismeasurement component of the productivity slowdown is the Denison figure of decline in quality of labor inputs of around 0.1 percentage point.

At the end of the day, I would have to conclude that there's not much gold in the hills of mismeasurement. But this conclusion was foreordained in Baily and Gordon's table 2, where we saw that the productivity slowdown was as pronounced in industries with well-measured output—mining, manufacturing (outside of computers), and utilities—as it was in poorly measured industries like services and FIRE. Those who toil in the vineyards of productivity are not yet out of work.

David Romer: In the good old days, productivity data were like most other standard economic time series: they were quite useful as long as you did not make the mistake of thinking hard about where they came from. In the case of price indexes, for example, if one were to set out to construct an economically appropriate price index, one would soon find that there were deep problems involving the absence of any representative consumer in the economy, the treatment of new commodities, the treatment of quality changes, and so on, that probably made the construction of a "valid" price index impossible. Still, knowing all that, economists rely on price indexes: for example, none of us seriously doubts that the indexes that show inflation at roughly 2 percent in the mid-1960s and at roughly 10 percent in 1980 reflect a genuine change in the economy. Similarly, despite the profound conceptual difficulties in measuring productivity, economists have generally believed that the measured slowdown in the growth of output per worker-hour after 1973 reflected a real and important change in the economy.

The central message that I take from the paper by Baily and Gordon is that the view that productivity data, despite whatever imperfections they may have, are a useful tool is simply not correct in the case of industry-level data. Baily and Gordon criticize disaggregate productivity data on two fronts. First, they point out several readily identifiable and quantitatively large problems with the data—such as the measurement of output by labor input in some industries and the inadequate or nonexistent treatment of changes in quality in others. Second, Baily and Gordon point out that the data are often obviously unreasonable. The most striking example of this is that measured output per worker-hour in construction is lower today than it was 40 years ago.

The long-term solution to the problem of inadequate industry productivity data is easy to see. It is to rebuild the data from the ground up—to do a serious job of measuring output by industry and to account systematically for changes in quality. However, it is an open question whether that is going to occur in the foreseeable future.

In the meantime, there is a possible short-term solution to the problems with the industry-level productivity data that Baily and Gordon appear to be sympathetic toward, but that I think should be treated with caution. That solution is to make corrections to the data until one obtains estimates that appear reasonable. For example, Baily and Gordon note that plausible corrections to the structures deflator could yield sensible figures for productivity growth in construction, and that sensible adjustments for the convenience associated with branch banking and for new financial instruments could result in a more reasonable banking productivity series. The problem with this approach is that when the data are fraught with difficulties, then if correcting one problem does not produce estimates that appear reasonable, it is easy to find additional problems to correct. The procedure can be continued until something approximating the desired figures is obtained. If one is not careful, this approach may not be dramatically different from simply making up the data.

In contrast to their findings concerning industry-level data, Baily and Gordon conclude that the measurement issues that they explore are probably not important to variations in measured aggregate productivity growth. I find this plausible, for two reasons. First, as Baily and Gordon point out, these issues and errors are longstanding; although they may cause trend productivity growth to be mismeasured, there is no particular reason to expect the size of the error to change over time, and thus no particular reason to expect measurement error to account for large variations in productivity growth rates. Second, most industries, particularly those where measurement problems seem especially severe, are small relative to the economy; even large changes in the error in measuring productivity growth in one or two industries will have little impact on measured aggregate productivity growth. Thus I come away from the paper with my initial belief concerning aggregate productivity statistics—that although they suffer from a variety of problems both in principle and in practice, they are a useful tool that for the most part reflects genuine economic developments—largely unshaken.

The paper thus makes two main contributions. The first is warning economists away from disaggregate productivity data: unless one is interested in correcting these data, it appears to be a good rule simply not to use them. It is a nice change to have a paper pointing out the pitfalls in a set of data before rather than after dozens of papers have been written drawing striking conclusions from those data. The paper's second contribution is its uncovering of numerous errors, inconsistencies, and questionable practices in the productivity data. Baily and Gordon's efforts will, I hope, be put to good use by the BEA and the BLS in improving industry and (to a smaller extent) aggregate productivity measurement, and by others working in the field.

Let me now turn to the "computer puzzle." One of the central questions running through the paper is "What have all those computers been doing?" or, more prosaically, "Why has the vast increase in investment in computer power not been reflected in higher measured productivity growth?" It seems to me that there is no mystery here at all. It is a basic rule of growth accounting that large changes in investment cause only small changes in output. The reasons for this are that investment is a small fraction of GNP and that the marginal product of capital is small. Since computers are a quite small part of total investment, a vast increase in investment in computers would yield only a small increase in measured output even if all the computers were being used productively and were generating measured output.

To be more precise about this, consider the following calculation. Suppose that computers depreciate linearly over eight years and that the marginal product of capital is 15 percent; reasonable variations in these parameters would have little effect on what follows. With these parameters, the stock of real computing capital grew by a factor of 30 from 1965 to 1986. Despite this vast increase, however, the stock of computing capital in 1986 amounted only to about $210 billion in 1982 dollars, or about 6 percent of a year's GNP.[1] If the marginal product of capital is 0.15, it follows that computers are increasing output by slightly under 1 percent. These calculations imply that if computers are being used productively, they have raised the average annual growth rate of output over the past two decades by roughly a twenty-fifth of a percentage point. I can imagine sensible variations on this calculation that would raise or lower this figure, either for the economy as a whole or for specific

1. I measure computer investment using NIPA, table 5.7, line 4—"Office, computing, and accounting machinery." The higher capital stock figures in Baily and Gordon's table 12 simply reflect the fact that their figures include communication equipment, instruments, and photocopy and related equipment.

industries, by a few factors of two. But the number seems to be in the right ballpark. In short, asking why the vast investment in computers has not had a discernible impact on productivity growth is a little like asking why the pull of gravity is not noticeably stronger when the moon is on the opposite side of the earth than it is when the moon is above us.

I would like to conclude with a more speculative comment about the relationship between what the productivity data tell us and popular perceptions of changes in standards of living. According to the data, productivity growth has slowed in recent decades but has continued to be positive. But popular perceptions appear to be quite different: the economic circumstances of ordinary Americans are widely perceived to be worse, not better, than they were two decades ago. For example, most Americans believe that the main reason that so many more women are working today than in the 1950s and 1960s is simply economic necessity. Now I have no doubt that the story told by the data is broadly correct: a moment's reflection is enough to convince one that the quantity and quality of goods that can be commanded by a typical worker's wages today are greatly superior to what they were a generation ago. But this raises a puzzle: why does the public appear to perceive otherwise? For example, is the *change* in the growth rate of productivity an important determinant of economic satisfaction? Or might it be that achieving a certain positive growth rate is needed simply to make us feel no worse off than before?

The purpose of economic growth is not to increase material wealth but to increase economic satisfaction and well-being. Answering questions like the ones above is thus extremely important. If it is true that while at any given time an increase in output would raise economic satisfaction but that in the long run there is no relation between the absolute level of income and economic satisfaction, then pursuing higher output is a futile way of attempting to reach the underlying goal of improved economic well-being. Since it is surely the case that productivity growth is not the only determinant of economic well-being, and since it may not even be the main determinant, perhaps a more pressing task than attempting to understand variations in productivity growth is trying to understand what in fact are the major determinants of economic well-being.

General Discussion

Robert Hall believed that the authors' investigation brought into focus five developments that can explain all but a small fraction of the slowdown in productivity growth: overcapacity and pollution controls in electric utilities, safety and environmental regulations in mining and the depletion of mineral resources, mismeasurement of airline discounts, increased product variety in retailing and elsewhere, and a shift toward remodeling and customized construction. He noted that a recent paper by Kevin M. Murphy, Andre Shleifer, and Robert Vishny provides a theoretical basis for the importance of the last two factors.[1] According to those authors, the growth of the middle class creates mass markets for relatively homogeneous goods that exploit scale economies in production and distribution. More recently, the opposite shift has occurred in the United States as the distribution of income has tilted in favor of the upper-income groups, who prefer to consume specialty goods produced on a much smaller scale.

Charles Schultze disagreed with Hall's emphasis on environmental policy for explaining the slowdown in productivity growth, noting that Edward Denison had found these effects to be small. Schultze also questioned Baily and Gordon's presumption that mismeasurement could possibly explain the productivity slowdown. Even if it could explain the slowdown in the United States, mismeasurement would not account for the simultaneous slowdown worldwide. Baily noted that the slowdown outside the United States could simply mean that the technology gap between the rest of the developed world and the United States had been closed. The contribution to productivity growth that came from closing that gap is no longer available to the rest of the world. Bradford De Long noted that the exhaustion of productivity growth has a precedent: it recalls the dramatic slowdown in the growth of Great Britain's GNP per capita in the first quarter of this century.

The discussion turned to the dramatic decline in the *level* of productivity in construction. Gregory Mankiw noted that unmeasured increases in the quality of housing coming from customization should explain not

1. Kevin M. Murphy, Andre Shleifer, and Robert Vishny, "Industrialization and the Big Push," *Journal of Political Economy* (forthcoming, 1989).

only the decline in measured construction productivity but a 20 percent to 30 percent increase in measured real housing prices as well. Because the price of new homes has not increased that much relative to the price of existing homes, Mankiw was skeptical of the customization explanation. Robert Gordon noted that much of the increase in the cost of housing reflects escalation of land prices, which is common to both new and old housing and masks the growth in construction costs.

Schultze pointed out that the comparison of U.S. and Canadian construction productivity and prices presented in table 12 deepens the puzzle. Over the comparison period of almost 40 years, the real price of structures in both countries fluctuated but showed no significant trend. Yet relative to national productivity growth, construction productivity performed much better in Canada than in the United States. If failure to capture quality improvements in the U.S. price index was the cause for the relatively poor U.S. productivity performance, the reported real construction price should have risen substantially faster in the United States than in Canada.

Jack Triplett reported on work at the Bureau of Economic Analysis using superlative index numbers to create a new series for real output. Compared with the current deflation procedure, the new output series grows faster before the benchmark year (1982) and slower afterwards. He also noted that the current BEA airline deflator averages the CPI index and an index of average revenue per passenger mile (on the grounds that the two measures have offsetting quality errors). Finally, Triplett felt that David Romer was overly skeptical about the value of industry-level productivity measures. Because BLS industry productivity data use output in the numerator, while the authors use as their numerator BEA gross product originating by industry (equivalent to value added), the two series will diverge at times, even if both are measured without error.

Michael Lovell agreed that increases in product variety are valuable but reasoned that some method of measuring their value was needed. To quantify the benefits to the consumer from increased quality and variety he referred to a revealed preference test invented by Richard Ruggles. Individuals would be given a choice of spending their income on items in a 1988 Sears catalog or some fraction of their income on items from the 1973 catalogue. The income fraction at which there is indifference would reflect the point at which changes in variety and quality offset prices. Finally, he noted that the socially optimal degree of product differentiation rises when a technology such as computers lowers setup costs.

Henry Aaron noted that test scores fell simultaneously for all grade levels, suggesting the fall may tell more about social attitudes than about the quality of education acquired by new entrants to the labor force. It may be true that labor quality has diminished at the same time that test scores have fallen, but this is not evidence of causation.