TECHNICAL WORKING PAPER SERIES

THE NBER MANUFACTURING PRODUCTIVITY DATABASE

Eric J. Bartelsman
Wayne Gray

Technical Working Paper 205

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1996

This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper 205
October 1996

THE NBER MANUFACTURING PRODUCTIVITY DATABASE

ABSTRACT

This paper provides technical documentation to accompany the NBER manufacturing productivity (MP) database. The database contains information on 450 4-digit manufacturing industries for the period 1958 through 1991. The data are compiled from various official sources, most notably the Annual Survey of Manufactures and Census of Manufactures. Also provided are estimates of total factor productivity (TFP) growth for each industry. The paper further discusses alternate methods of deflation and aggregation and their impact on TFP calculations.

Eric J. Bartelsman
Central Planning Bureau
Van Stolkweg 14
Postbus 80510
2508 GM The Hague
THE NETHERLANDS

Wayne Gray
Department of Economics
Clark University
950 Main Street
Worcester, MA 01610
and NBER
WGRAY@VAX.CLARKU.EDU

1 Introduction

The NBER Manufacturing Productivity (MP) database contains annual information on 450 manufacturing industries from 1958 to 1991. The industries are those defined in the 1972 Standard Industrial Classification, and cover the entire manufacturing sector. The data themselves come from various government data sources, with many of the variables taken directly from the Census Bureau's Annual Survey of Manufactures and Census of Manufactures. The advantages of using the MP database are that it gathers together many years of data, adjust for changes in industry definitions over time, and links in a few additional key variables (i.e. price deflators and capital stocks).

The initial version of the data was developed for the Census Bureau as a joint project by researchers from the University of Pennsylvania and SRI International.[1] The data was developed to compare survey-based capacity utilization data with estimated capacity utilization based on production function estimation. This required a relatively long time series of consistent industry data in real terms, along with industry capital stocks.

The initial MP data has been updated several times in recent years. From 1980 through 1986, revisions were done by Wayne Gray at NBER. These revisions, besides extending the data, provided improvements in the energy deflators allowing for more variation across industries in the movements of energy prices. There were also improvements in the output price deflators, as the BLS moved from a product-based deflator to an industry-based deflator. The most recent update, carrying the data through 1991, was primarily carried out by Eric Bartelsman at the Federal Reserve Board. It takes advantage of new data for constructing the capital stocks and deflators; consequently some variables have been revised. The real capital stocks were also shifted to a benchmark year of 1987 (rather than 1972), while the deflators use benchmark information from 1972, 1977, and 1987.

The MP database has been used in a wide variety of research projects. The initial purpose often was estimation of production functions using industry-level data (e.g. Bartelsman (1995) and Bartelsman et al. (1994)) This research strategy has been in part supplanted by the availability of plant-level productivity data. However, the industry-specific price deflators have been used in a variety of plant-level productivity analyses since plant-specific deflators are generally not available (e.g. Bartelsman and Dhrymes (1994), Baily et al. (1992) and (1996b). Some work in labor economics has been done using the information on em-

[1]This discussion is based primarily on the Andrews and Zabala (1984) documentation, along with a data appendix from Crawford, et al. (1980).

ployment and wages linked to other industry characteristics (e.g. Dunne and Schmitz (1995) and Berman et al. (Berman, Bound and Griliches 1994)). The data have also been used to calculate industry productivity growth rates, using growth accounting methods. These industry productivity measures have been used as dependent variables, looking at a variety of possible influences on productivity, including various measures of government regulation (e.g. Gray (1987)). They have also been used as independent variables in other cross-industry analyses.

The MP database is now available on the Internet, at nber.harvard.edu, in the directory /pub/productivity. It can be accessed using anonymous ftp.[2] At the moment the directory contains 7 files: readme.doc (explaining how to put the data together), nberprod.doc (describing the data), asm1.dat and asm2.dat (each holding half the data - see readme.doc for information), asm1.dat.Z and asm2.dat.Z (compressed data files, using Unix compression, to reduce file transfer time), and asm12.zip (contains asm1 and asm2, compressed using PKZIP for DOS computers).

The MP database can also be accessed through the World Wide Web, through the NBER Home Page at http://nber.harvard.edu, under the 'Online Data' category. Alternatively, the URL for the MP data directory is http://nber.harvard.edu/pub/productivity, which enables you to skip two pages of menu selections. You have the same choice of files to download as described above (readme.doc and nberprod.doc to describe the data; asm1.dat+asm2.dat, asm1.dat.Z+asm2.dat.Z, or asm12.zip which contain the data).

The MP database is likely to be updated as more data become available or as problems with the data are pointed out and corrected. For this reason, we ask that you refer people to the original Internet source rather than having you provide them a copy of your version of the data (especially if your copy is months or years old). If you are using the data often, you should probably also occasionally check the original Internet files to ensure that you haven't missed some data improvements.

Anyone wanting to be informed by e-mail of changes to the MP database should send a brief request to wgray@vax.clarku.edu; your name will be added to a mailing list. Please send questions/comments/problem reports to the same address, and we will try to respond to them (both directly and by correcting problems with the database as they are discovered). Feel free to use the data in your research, with appropriate attribution (a reference to this

[2]To access the files: 1) run your ftp program to connect to nber.harvard.edu. 2) login as 'anonymous' with the password being your internet account (e.g. 'john@harvard.edu'). 3) change directories - 'cd /pub/productivity'. 4) get files - 'get *fname*'.

NBER Technical Working Paper). Please let others know about its existence. We'd like to think that the effort involved in assembling the data will pay off in improved research opportunities.

Section 2 describes the variables included in the dataset and the sources from which they were constructed. Section 3 contains a discussion of some of the conceptual (and practical) problems involved in working with the data, especially for doing productivity analyses. Section 4 presents some descriptive views of the data and analyzes the impact of changing various assumptions when calculating industry productivity growth rates or the aggregate productivity growth rate for the manufacturing sector.

2 Data Sources

2.1 ASM/Census Data

Most of the variables in the database come from the Annual Survey of Manufactures (ASM). The ASM is a sample of around 60,000 manufacturing establishments, carried out by the Census Bureau. The sample is drawn from the Census of Manufactures (CM), done every five years, which is designed to collect data on all of the manufacturing establishments in the country.[3] The ASM and CM both collect data on a relatively stable set of basic variables. The CM also collects data on a changing set of additional variables (e.g. computer hardware and software expenditures in the 1987 CM). The MP database contains only the variables available in the ASM, although it is certainly possible to link in CM industry data (as many researchers have done).

The basic information in the ASM is used for eleven of the eighteen variables in the data set. These are number of workers, total payroll, number of production workers, number of production worker hours, total production worker wages, value of shipments, value added, end-of-year inventories, new capital investment, expenditure on energy, and expenditure on materials (including energy). All of these variables are in millions of (nominal) dollars, except for the labor input variables which are in thousands of workers and millions of worker hours. A more complete description of the variables is given in the data appendix.

[3]Some of the smallest plants covered by the Census have their data imputed from administrative records, to reduce the record- keeping burden. Also, there was a switch in the timing of the CM during the 1960s, so the Census years in our data are 1958, 1963, 1967, 1972, 1977, 1982, and 1987.

2.2 Industry Capital Stocks and Investment Deflators

The original ASM data contains information on new investment spending, but does not measure the total capital stock for the industry. The CM data includes information on the gross book value of the capital stock, but these measures may be influenced by the accounting methods used to calculate book value. One of the major efforts of the original Penn-SRI research project was to create real industry capital stocks as described in Crawford et al. (1980).

As in previous revisions of the MP database, the starting point for the process of creating real capital stock series was a set of three-digit industry capital stock estimates. In earlier versions, these originated at the Commerce Department, and were based on investment flow and book value data extending back to 1947 or earlier. In the present revision, we use FRB net capital stocks as the basis of our 4-digit estimates.

The FRB 3-digit net capital stock data are based on 3-digit investment series for plant and equipment (from ASM), the 1977 industry-asset type investment flow matrix, producer durable equipment deflators, and a table of mean service lives by asset type (from BEA). The first step is to use bi-proportional balancing to extrapolate the investment flow matrix from 1890 through 1991. The appropriate asset deflator is then applied to industry investment by type (28 types are used in manufacturing) to get real investment by industry and type in 1987 dollars. Using a perpetual inventory model, with stochastic service lives and beta decay, provides us with the 3-digit real net capital stocks. For a more complete description of this project see Mohr (1995).

The 3-digit data are converted to the 4-digit level by assuming that the industry-asset type flows are the same for all 4-digit industries within a 3-digit. With this information, 4-digit investment deflators are created for equipment and structures separately, and initial 4-digit real capital stocks for 1958 are created using the ratio of 4-digit to 3-digit net capital from the original Penn-Census-SRI data. Using the implied "depreciation" series from the 3-digit system, applied to each appropriate 4-digit, we can successively add real equipment and structures investment and subtract the "depreciation" to create real net capital stocks from 1958 through 1991.

Some summary of information about the investment flow data and the resulting growth rates of the capital stocks are given in tables 1 and 2. The second and third columns of table 1 give the share of investment going to equipment for each 2-digit industry in 1977 and 1987, respectively. The next two columns show computer investment as a share of equipment investment. Whereas equipment shares have been rising only slightly, and do not vary all that much across industries, the computer share has essentially doubled, and shows sizable variation across industries.

Table 2 shows growth rates of the real capital stock for each 2-digit industry for selected periods. The growth rates are compared with the figures which had been available in the previous releases of the MP database. Owing to disaggregation into more detailed asset types, specifically for computer equipment, it is not surprising that the results vary so much. In general, capital stock growth rates for the 1982–1987 period are lower using the updated methodology.

2.3 Calculation of Deflators

Prior to 1972, the output deflator (actually, a value of shipments deflator) comes from the BEA, and is itself based on detailed producer price indices from the Bureau of Labor Statistics, supplemented by a few specialized deflators for military goods from the government division of BEA. Their general procedure in the past was to find an appropriate price index for each 7-digit product within each industry, then weight each one by the share of that product in industry production. As the available price indices have changed over time, there have been some changes in the indices used but there is no comprehensive record of which indices were used for the years prior to 1972.

From 1972 on we have used 5-digit product deflators from BEA. These are largely created from BLS's industry-based producer prices which are extrapolated backwards using the old BLS product prices. We then use the "Make" tables (describing production of 5-digit product by 4-digit industry) for 1972, 77, and 87, to create 4-digit output deflators. Rather than relying on one benchmark year, we have created "benchmark years weighted" deflators, which are Fischer Ideal indexes of the individual fixed-weight deflators.

Besides the BLS source data, the 5-digit price data from BEA contains their computer deflator, which is adjusted for quality change using hedonic techniques. The same source material for this deflator is also used for the deflator for computer equipment used in the computation of real investment and real capital stocks.

Output growth rates are very sensitive to the use of hedonics, and the choice of index number methodology. For long-run productivity comparisons, fixed-weight price indexes can substantially bias results, especially if underlying price trends are divergent. Further, choices made regarding the output deflators need to be made consistently for the creation of capital stocks and materials prices in order to understand the effects on TFP calculations. Section

4 discusses these issues and provides alternate calculations of TFP growth.

The materials deflators were created by averaging together price deflators for 529 inputs (369 manufacturing industries and 160 non-manufacturing industries), using as weights the relative size of each industry's purchases of that input in the Census Department's Input-Output Tables. The tables were obtained for 1972, 1977, 1982, and 1987, with pre-1972 data given 1972 weights, post-1987 data given 1987 weights, and interpolation between the I-O tables used for the remaining years. The shipments deflators for each of the 4-digit SIC manufacturing industries were used for the manufactured inputs. Non-manufacturing prices (agriculture and mining) were taken from the corresponding BLS sector's price indexes. The inflation in materials prices was calculated as a Divisia index, with each product's inflation rate weighted by the average of previous and current-year's shares in total materials used. A comparison of the old and new materials deflators is given in table 3. The new deflators show substantially lower price growth during the 1980s, most notably for petroleum and chemicals.

The energy price deflator is based on each industry's expenditures on six types of energy (electricity, residual fuel oil, distillates, coal, coke, and natural gas). These six types of energy made up 94.6 percent of all energy expenditures by the manufacturing sector in 1976. The growth (or decline) in price for each energy type was calculated on whatever level of industry detail was available. We use the National Energy Accounts database for 1958-85 and the Energy Department's State Energy Price and Expenditure Report for 1986-91. These individual price changes were weighted by expenditure shares to create a Divisia index of energy prices. Energy expenditures for individual 4-digit industries are available only from the Annual Survey of Manufactures for 1974-82. Earlier years are taken from the National Energy Accounts and later years are taken from the 1985 and 1988 Manufacturing Energy Consumption Surveys, which have less industry detail.

2.4 Productivity Measures

A measure of total factor productivity growth has also been included in the MP dataset. It is based on a five-factor production function: capital, production worker hours, non-production workers, non- energy materials, and energy. A Divisia index of TFP growth is calculated as the growth rate of output (real shipments) minus the revenue-share-weighted average of the growth rates of each of the five inputs. The shares (average of current and previous period) are taken from the ASM data on the expenditures for each input, divided by the industry's value of shipments. Capital's share is calculated as a residual, so the shares add to 1. The labor inputs are measured in real terms as the number of production worker hours and number of non-production workers. Nominal expenditures on energy and non-energy materials are deflated by their respective deflators. The real capital input is assumed to be proportional to the measured real capital stock, so the capital stock growth rate is used to measure capital input growth; there is no adjustment for capacity utilization.

3 Complications

3.1 Redefinition of Industries

The use of the 1972 SIC for our industry definitions comes about because the initial MP data extended only until 1976, so the 1972 industry definitions were the most recent ones available. One major problem in creating any long-term industry data set is that industrial classifications have changed over time to reflect changes in the importance of different industries in the economy. These changes have occurred to a small degree in every Census of Manufactures, and to a much greater extent every few Censuses, with major redefinitions in 1958, 1972, and 1987. In order to look at changes in the economy, and to measure growth rates over time, it is important that the industry definitions be kept as consistent as possible. Therefore the published industry data from some years must be re-allocated to fit an earlier (or later) set of industry definitions.

One possible route would be to use plant-level data such as the Longitudinal Research Database (LRD) maintained at the Census Bureau's Center for Economic Studies. Each plant could be assigned an appropriate industry code (from whatever vintage of SIC codes is desired) and followed throughout the industry definition changes. There would still be some theoretical issues (even within a plant the mix of products produced might change, and it might be difficult in some cases to map data based on one set of product classifications into different industry classifications), but there are more practical considerations that rule out this approach. First, the LRD's annual data extend back only to 1972, so the dataset would lose 14 years of information. Second, making sure that the data for redefined industries met the Census Bureau's standards for disclosure would be very difficult, since the differences between the published and the redefined industry data might be attributable to a small number of plants with changed industry definitions.

Instead, we rely on the published industry data and re-allocate the data for industries which changed definition during the period. In the years when major re-definitions occurred, the Census publishes 'concordances', identifying the fractions of each former industry that go into each new industry (and vice versa). In the original MP data, all of the information from 1958 to 1971 was apportioned to the 1972 industries based on their value added in 1972.

For example, the 1967 SIC industry 2015 (poultry dressing plants) was split into two 1972 SIC industries, 2016 (poultry dressing plants) and 2017 (poultry and egg processing). In 1972 the establishments formerly classified in SIC 2015 had total value added of $892.9 million. Of that total, $724.4 million was in establishments which were classified as SIC 2016, while $168.5 was in establishments classified as SIC 2017. Thus SIC 2016 represents 81 percent of SIC 2015, and for the 1958-71 period each published variable for SIC 2015 would have 81 percent of its value allocated to SIC 2016 and 19 percent allocated to SIC 2017.

Some of the industry re-definitions are more complicated, with pieces of three or more industries being split and joined to form new industries, but the basic procedure remains the same: identify the fraction that each 1972 SIC industry represents of the other SIC definitions, multiply the published value by that fraction, and add up the pieces. In the 1987 SIC revisions, the Census Bureau published concordances for several of the basic variables (not just value added), so we could use separate mappings for each variable.[4]

The main concern with using this sort of fixed-weight allocation method is the likelihood that the mix of production in the economy changes over time. In the example described earlier, the concordance showed that 19 percent of the 1958 SIC industry 'poultry dressing plants' was classified as 'poultry and egg processing' under the 1972 SIC codes, but this reflects the 1972 mix of production. It is quite likely that in 1958 the mix of production was different—perhaps markedly so—but we are stuck using the 1972 concordance for allocating the entire 1958-1971 period. This will tend to understate the growth (or decline) of industries over time, since the different segments of each 1958 industry are assumed to grow at the same rate for the 1958-1971 period. This effect is exacerbated because industry redefinitions tend to occur in sectors where new products are growing (or old products are shrinking).[5]

[4]The variables with concordances were EMP, PAY, PRODE, PRODH, PRODW, VADD, MATCOST, VSHIP, INVENT, and INVEST. In the few cases where one concordance value was missing for a particular industry, we used the most closely related variable that was present for that industry. There was no concordance for the ENERGY variable, so the MATCOST concordance was used instead.

[5]When a few more years of LRD data have accumulated, it may be possible to test for the extent of this bias in the mapping from 1987 SIC codes into 1972 SIC codes, by looking at the plant-level data within industries that were redefined, to see whether the fixed weights understate the growing industries and overstate

A minor redefinition also occurred in 1977, with SICs 2793 and 2794 combined to form 2794, SICs 3671, 3672 and 3673 combined to form 3671, and SIC 3713 split into SICs 3713 and 3716. The same procedure described above was used here, with the value added of each industry in 1976 used to allocate the new data for SIC 2794 and 3671 among the respective 1972 SIC industries. The new data for SICs 3713 and 3716 was combined to maintain comparability with the old SIC 3713 data. These changes maintained the number of industries at 450. (although the actual number of industries in the 1977-1986 ASM/CM data is only 448).

3.2 Missing Data

A second issue is the imputation of missing data values. When there are only a few establishments in a particular industry the Census Bureau will sometimes not report data for that industry, to avoid disclosing data for an individual establishment. This is quite rare in years when the Census of Manufactures was done and in recent years, but is fairly common in the earlier years of the Annual Survey of Manufactures. It is not entirely clear what procedure was used in the initial MP dataset to impute these values.

The procedure used in this dataset takes advantage of the existence of 2- and 3-digit industry data. The 2-digit data was available for virtually all cases (and was imputed, based on values of that variable for surrounding years and for related variables in that year, when it was missing). The complete 2-digit data and available 3-digit data were used to fill in missing 3-digit variables. The 3-digit industries within that 2-digit industry which had that variable present were summed and subtracted from the 2-digit industry's value to calculate the 2-digit remainder. This remainder was then allocated among all 3-digit industries with missing values, based on their respective values for that variable in surrounding years. This procedure was repeated to fill in missing 4-digit industry data, based on the (now complete) 3-digit data.

3.3 Further Notes on Data Idiosyncracies

There are a number of points about the data that should be mentioned, to avoid confusion about what is being measured. For example, the value given in the database for materials expenditure includes expenditures on fuels and electrical energy. Therefore, to separately

the declining ones, as expected, and how large this bias is.

identify the expenditure on non-energy materials, you need to subtract the energy expenditure from the materials expenditure. To get the real value of non-energy materials you should first deflate all materials, then subtract off the deflated value of energy materials.

The measure of industry output used here is actually a sales measure, not a production measure (value of shipments, not adjusted for changes in inventories). One could try to adjust this for inventory change (using the end-of-year inventories variable), but there are problems with this: before 1982 the data can refer to any generally accepted accounting method, and there could be problems when the accounting method was changed; also, this dataset does not disaggregate inventories by stage of production process, so it's not obvious which deflator should be used to calculate real inventory change.

The measures of the number of employees exclude those employees in auxiliary units (headquarters and support facilities), and the measures of payroll exclude their wages. This is a fairly sizable, and growing, segment of the workforce (994,300 in 1972 and 1,283,200 in 1986; 5.2 and 7.0 percent of total manufacturing employment; accounting for 7.9 and 10.7 percent of total payroll in manufacturing, respectively), but they are not allocated to individual industries in the published data. They could be allocated among these industries, based on a breakdown of auxiliary employment by two-digit industry published in the Census years, but this has not been done here.

The measures of payroll (both for all employees and production workers) do not include Social Security contributions or other legally mandated payments, and may exclude employer payments for some fringe benefits. In recent years, this data has been published in supplemental tables to the ASM data, and so it could (in principle) be added to the dataset. For example, in 1986 these supplemental labor costs made up $92.3 billion, or 18.6 percent of total compensation.

4 Aggregate TFP growth - sensitivity analysis

The TFP variable for each industry included in the MP database is constructed in the following manner:

$$\hat{\text{TFP}} = \hat{Q} - \sum_i \alpha_i \hat{X}_i, i \in \text{K,N,L,M,E}$$

where Q is real output, α_i is the share (average of current and lagged year) in revenue of factor i, X_i is the real input of factor i, and a $\hat{}$ denotes a log first difference. The share of capital is computed as one minus the sum of the other shares.

Aggregate TFP is calculated as:

$$\hat{\text{TFP}} = \sum_j \phi_j \hat{\text{TFP}}_j$$

where j indexes industries, and ϕ_j is an appropriate weight for each 4-digit industry.

The TFP measure turns out to be very sensitive to the types of deflators applied to output and materials, and aggregate TFP growth measures are sensitive to the choice of ϕ. In this section, the influence of different methods of deflation and aggregation on TFP is assessed.

4.1 Deflators

One of the most important effects on TFP growth comes from assumptions regarding incorporation of quality change in the deflator. In the MP dataset, we incorporate the BEA hedonic deflators for the computer industry. The deflators are used consistently for the computation of the output deflators, materials deflators, and computation of the real capital stock. The computer components for which the BEA computes hedonic deflators are shown in the first six rows of table 4. These components are then matched to the 5-digit products produced primarily in the 4-digit computer industry (3573). The total computer deflator is also used to deflate the computer investments used in creation of the capital stocks. The output price for SIC 3573 which is included in the MP database is a benchmark years weighted index (Fisher Ideal) of the 5-digit product deflators, and is denoted by 'BW' in table 4.

As an alternative, the total computer deflator from BEA is used instead of the BLS deflators from 1977 onward for an additional four 5-digit products such as calculators, office copiers, and telephone switching equipment (codes 35792, 36741, 36749, 36611). The resulting 5-digit deflators are then also weighted using the industry by product make tables from 72, 82, and 87, and are denoted by 'Hed' in the table. Next, the deflator for calculators is used for all the components of the computer industry, turning off the hedonic methods used by BEA. This is labelled by 'NoH' in the table. The next row in the table uses only the 1987 4-digit industry by 5-digit product make table to aggregate the 5-digit deflators, and is labelled 'F87'. Finally, the deflators of the previous release of the dataset, which was computed with the 1972 make table, is labelled 'Gray.' The next three rows of the table show the resulting deflators for the three 4-digit industries which are affected by the addition of hedonics to the above-mentioned 5-digit products. The following section of rows shows the average annual price changes for total manufacturing output as Divisia weighted aggregates of the 4-digit deflators as described above. Finally, the last two rows show how different aggregations, fixed 1987 weights and implicit weights, differ from the Divisia weighted aggregate of the 4-digit BW deflators. A similar pattern of upward bias post weight-year and lower growth prior to weight-year for the fix weighted aggregation occurs for all the variants of the 4-digit deflators.

In the first three rows of table 5 and in figure 1, the effect of the deflator assumptions on aggregate TFP can be seen. Figure 1 shows that the 'Hed' TFP measures grows faster than the 'BW' measure, while the 'NoH' measure has the slowest growth, as should be expected.

Another issue with regard to the deflators are how the 5-digit product deflators are aggregated to the 4-digit industry level. The output price series—named PISHIP in the MP dataset—are 'benchmark years weighted' deflators. This concept borrows its name from the experimental GDP measures published by BEA, prior to BEA's adoption of chain weighted GDP. The benchmark-years weighted deflators are computed as a geometric mean of two fixed weight deflators. For example, for the 1982 through 1987 period, the deflator is the geometric mean of the 4-digit deflator based on the 1982 make table weights and the 4-digit deflator based on the 1987 make table weights. For the years 1972 through 1982, the deflator is a geometric mean based on 1972 and 1982 weights.[6] Prior to 1972, the deflators are based on 1972 make table weights.

Figure 2 shows the movements in aggregate manufacturing TFP using different 4-digit shipments deflators. The solid line shows the benchmark years weighted measure, the long dash—labeled BEA in the figure—is based on fixed 1987 weights, and the dotted line—labeled GR—uses fixed 1982 weights. Fixed weighting understates productivity growth in periods before the base year, and overstates growth in later years. The growth in the fixed 1982 based measure is much faster in the years after 1982 than the benchmark years weighted measure.

Using different hedonic schemes and different 4-digit deflator concepts affects TFP calculations through both real output growth and real input growth, through the consistent calculation of materials prices and capital stocks. For this reason, the ultimate effect of a change in deflator assumption on TFP growth cannot be ascertained a-priori: a reduction in the growth rate of computer prices increases output in that industry, but feeds through to

[6]The 1972 weighted data had no hedonics for SIC 3573. For this industry all data prior to 1982 are based on the 1982 make table weights.

increased materials usage and differences in capital growth in other industries. The net effect of these changes on aggregate manufacturing TFP depend on the method of aggregating industry TFP, described in the next section.

4.2 Aggregation

Industry TFP growth is measured as the percentage increase in gross output less the percentage increase in (weighted) capital, labor, and materials inputs. Ideally, the measure of TFP for manufacturing as a whole would have a similar interpretation: the increase in gross output being shipped to sectors outside of manufacturing less the increase in, appropriately weighted, capital, labor, and materials from sectors outside manufacturing. Domar (1961) provided the framework such an aggregation scheme, which is now known as net-output, or Domar, weighting. The following example is based on Kortum (1995).

Consider a manufacturing sector with only two industries where the output of industry Y is shipped to the final goods sector, and the output of industry X is all used as material input in industry Y. Industry X uses materials, say live chickens denoted as M_x, from outside the manufacturing sector.

$$Y = A_y F(L_y, X), \text{and}$$

$$X = A_x G(L_x, M_x)$$

Growth in technology in each sector is given by the percentage change in A_i, and given constant returns to scale and perfect competition can be approximated with a TFP divisia index:

$$\hat{\text{TFP}}_y = \hat{y} - s_y \hat{l}_y - (1 - s_y)\hat{x}, \text{and}$$

$$\hat{\text{TFP}}_x = \hat{x} - s_x \hat{l}_x - (1 - s_x)\hat{m}_x$$

where s_i is the labor share of revenue in sector i, and $\hat{x} = \frac{\dot{X}}{X}$.

In order to derive some measure of aggregate TFP, various schemes have been used. One is to weight sectoral TFP growth by the sector's share in revenue. Another would be to weight with value added, in an attempt to avoid double counting of materials. The proposal by Domar to use net output weights, however, can be shown to be the same as TFP growth derived from a consolidated manufacturing sector. This is shown below:

Rewriting the final goods shipments as a function of primary inputs gives:

$$Y = A_y F(L_y, A_x G(L_x, M_x))$$

Given optimal allocation of resources, perfect competition, and constant returns to scale, we can derive:

$$\hat{y} = \hat{a_y} + (1 - s_y)\hat{a_x} + (s_y + s_x(1 - s_y))\hat{l} + (1 - s_y)(1 - s_x)\hat{m_x}$$

where s_i is the output elasticity of labor in industry i and optimal resource allocation ensures that L_y and L_x grow at the same rate. TFP growth for the net output of the manufacturing sector (which in this example is equal to Y) is given by

$$\hat{\text{TFP}} = \hat{a_y} + (1 - s_y)\hat{a_x} \tag{1}$$

To see that this is the same result given by net output weighting, consider the following input-output representation of the above example:

	Y	X	Ag	FD
Y				$P_y Y$
X	$P_x X$			
Ag		$P_m M_x$		
VA	L_y	L_x	L_{ag}	

For purposes of computing a TFP aggregate of sectors X and Y, we need to weight TFP growth in sector Y with the ratio of net output of sector Y to the net output of the combined sector X and Y. The net output of Y is $P_y Y$, which in our case also happens to be the net output of the combined sectors. The net output of sector X is the gross output minus what sector X supplies to itself. In our example this is just $P_x X$. The weight for sector Y TFP growth is therefore unity, while for sector X it is $\frac{P_x X}{P_y Y}$, which given optimal allocation of factor X in the production of Y is equal to $(1 - s_y)$, or one minus the labor share in sector Y—the same as equation 1.

The following table shows the weights given to sectoral TFP in creating an aggregate for sectors X and Y using different weighting schemes:

	Domar (net output)	Shipments	Value Added
Sector$Y$	1	$\frac{1}{2-s_y}$	$\frac{s_y}{s_y+s_x-s_y s_x}$
Sector$X$	$1-s_y$	$\frac{1-s_y}{2-s_y}$	$\frac{s_x-s_y s_x}{s_y+s_x-s_y s_x}$

Table 5 displays average TFP growth rates for the various 4-digit deflator types aggregated up to total manufacturing using Domar weights, time-varying value added weights (divisia), and 1987 fixed value added weights.[7] For each aggregation scheme, results are shown for the various deflator assumptions. As seen, no general statement can be made as to which aggregation scheme gives higher or lower TFP growth; the results depend crucially on the behavior of the disaggregated industries TFP growth. Table 6 shows the TFP growth rates for two-digit industries for each of the different deflator types, giving an impression of the differences in growth across sectors.

5 Conclusions

The foregoing should be considered a working document to be used in conjunction with the NBER Manufacturing Productivity Database. Over time, changes will be made to the document as the database is updated, new years are added, methodology is changed and SIC redefinitions take place. One major revision to be expected soon is the inclusion of a Thornquist Divisia index of capital service flows, based on appropriate weighting of component asset type stocks. In conjunction with this revision, a user cost of capital variable will be added to the database.

The sections showing alternative TFP results for various assumptions about deflators and aggregation schemes should be seen as illustrative of how sensitive aggregate TFP figures can be. For example, the switch from a fixed weight to varying weight deflation method greatly changes average growth rates prior and post the base year of the fixed weight scheme. Our choice of benchmark years weighting reflected experimentation with this type of deflation by the BEA for the NIPA. However, in the 1995 NIPA revision, the decision was made to use chain weighting instead. We will likely follow suit in an update of the database. Domar weighting should be considered the preferred method of aggregation, although value added

[7] Aggregates using shipments weights are not shown in the table.

weighting may provide a quick and dirty proxy. Obviously, the larger is the dispersion of sectoral TFP growth rates, the more sensitive is the aggregate to choices made for weights.

Anyone wanting to be informed by e-mail of changes to the MP database should send a brief request to wgray@vax.clarku.edu; your name will be added to a mailing list. Please send questions/comments/problem reports to the same address, and we will try to respond to them (both directly and by correcting problems with the database as they are discovered). Feel free to use the data in your research, with appropriate attribution (a reference to this NBER Technical Working Paper). Please let others know about its existence. We'd like to think that the effort involved in assembling the data will pay off in improved research opportunities.

References

Andrews, Steven and Craig Zabala, "Documentation of the SRI-Penn Manufacturing Industry Dataset," Technical Notes, U.S. Dept. of Commerce, Bureau of the Census 1984.

Baily, Martin N., Charles Hulten, and David Campbell, "The Distribution of Productivity in Manufacturing Plants," in "Brookings Papers: Microeconomics" Washington, D.C. 1992.

—— , **Eric J. Bartelsman, and John Haltiwanger**, "Downsizing and Productivity Growth: Myth or Reality?," *Small Business Economics*, August 1996, *8* (4), —.

—— , —— , **and** —— , "Labor Productivity: Structural Change and Cyclical Dynamics," NBER Working Paper Series 5503 March 1996.

Bartelsman, Eric J., "Of Empty Boxes: Returns to Scale Revisited," *Economics Letters*, July 1995, *49* (1), 59–67.

—— **and Phoebus J. Dhrymes**, "Productivity Dynamics: U.S. Manufacturing Plants 1972-1986," FEDS 94-1, Federal Reserve Board January 1994.

—— , **Ricardo J. Caballero, and Richard K. Lyons**, "Customer and Supplier Driven Externalies," *American Economic Review*, September 1994, *34* (84-4), 1075–84.

Berman, E., J. Bound, and Z. Griliches, "Changes in the Demand for Skilled Labor in US Manufacturing Industries: Evidence from the Annual Survey of Manufactures," *Quarterly Journal of Economics*, May 1994, *109* (5), 367–98.

Crawford, David, Gary R. Fromm, Lawrence R. Klein, and Frank Ripley, "The Census-SRI-Penn Data Set," 1980. draft.

Domar, Evsey D., "On the Measurement of Technological Change," *Economic Journal*, December 1961, *71* (284), 709–29.

Dunne, Timothy and James Schmitz Jr., "Wages, Employment Structure and Employer-Size Wage Premia: Their Relationship to Advanced-Technology Usage at U.S. Manufacturing Establishments," *Economica*, February 1995, *62* (1), 89–107.

Fromm, Gary, Lawrence R. Klein, Frank R. Ripley, and David Crawford, "Production Function Estimation of Capacity Utilization," 1980. data appendix.

Gollop, Frank M., "Modelling Aggregate Productivity Growth: The Impotrance of Intersectoral Transfer Prices and International Trade," *Review of Income and Wealth*, June 1987, *33* (2), 211–27.

Gray, Wayne B., "The Cost of Regulation: OSHA, EPA, and the Productivity Slowdown," *American Economic Review*, September 1987, *77* (4), 998–1006.

Gullickson, William, "Multifactor Productivity in Manufacturing Industries," *Monthly Labor Review*, October 1992, pp. 20–32.

Hulten, Charles, "Growth Accounting with Intermediate Inputs," *Review of Economic Studies*, 1978, *45* (3), 511–518.

Kortum, Sam and Simone Peart, "Aggregating Total Factor Productivity," June 1995. FRB memo.

Mohr, Mike F., "FRB Capital Stock Database," 1995. draft.

Data Appendix
Variable Descriptions and Comments

- **SIC, YEAR** - identify each observation in the dataset (SIC ranges from 2011 to 3999 and YEAR ranges from 58 to 91). The SIC refers to the 1972 industry classification system.

- **EMP** - number of employees (in 1,000s). This includes both production workers (average of reported employment in March, May, August, and November) and other workers (only March figure). It does not include employees in auxiliary (administrative) units, which are reported separately.

- **PAY** - total payroll (millions of dollars). This does not include social security or other legally mandated payments, or employer payments for some fringe benefits.

- **PRODE** - number of production workers (in 1,000s). This excludes supervisors above the line-supervisor level, clerical, sales, office, professional, and technical workers. Plants report the number of production workers for four separate pay periods (including the 12th day of March, May, August, and November), and these are averaged together.

- **PRODH** - number of production worker hours (in millions of hours). This includes all hours worked or paid for, including actual overtime hours, but excluding vacation, holidays, or sick leave.

- **PRODW** - production worker wages (millions of dollars).

- **VADD** - value added by manufacture (millions of dollars). This equals VSHIP - MATCOST + change in finished goods and work-in-process inventories during the year.

- **MATCOST** - cost of materials (millions of dollars). This includes the total delivered cost of raw materials, parts, and supplies put into production or used for repair and maintenance, along with purchased electric energy and fuels consumed for heat and power, and contract work done by others for the plant. This excludes the costs of services used, overhead costs, or expenditures related to plant expansion. Because MATCOST includes energy spending, to calculate spending on non-energy materials one must use (MATCOST-ENERGY).

- **ENERGY** - expenditures on purchased fuels and electrical energy (millions of dollars).
- **VSHIP** - value of industry shipments (millions of dollars). These are based on net selling values, f.o.b. plant, after discounts and allowances. This includes receipts for contract work and miscellaneous services provided by the plant to others. VSHIP is not adjusted for inventory changes, and can include large amounts of duplication across industries or even across plants in the same industry (especially machinery and transportation industries that often include production of parts as well as final products).
- **INVENT** - end-of-year inventories (millions of dollars). Before 1982 this was based on any generally accepted accounting method. Beginning in 1982 this is based either at cost or at market, with LIFO users asked to report pre-adjustment values.
- **INVEST** - new capital spending (millions of dollars). This includes permanent additions and major alterations to the plant structures along with new machinery and equipment. It combines spending on structures and equipment, and does not include used plant and equipment, land, or maintenance or repair expenses.

The following variables are not included directly in the ASM data, and their construction is described in the documentation above.

- **CAP** - real capital stock (millions of 1987 dollars). This equals (EQUIP + PLANT).
- **EQUIP** - real equipment capital stock (millions of 1987 dollars).
- **PLANT** - real structures capital stock (millions of 1987 dollars).
- **PISHIP** - price deflator for value of shipments (equals 1 in 1987).
- **PIMAT** - price deflator for materials (equals 1 in 1987). To match MATCOST, this is a deflator for all materials, including energy. If you need a separate 'non-energy materials' deflator, you can construct an implicit one, based on (MATCOST/PIMAT)-(ENERGY/PIEN).
- **PIEN** - price deflator for energy (1 in 1987).
- **PIINV** - price deflator for new investment (1 in 1987). This combines separate deflators for structures and equipment, based on the distribution of each type of asset in

the industry. Note that this is a deflator for new investment flows, not for the existing capital stock.

- **TFP** - five-factor total factor productivity growth (calculated from other variables as described in the documentation above, and expressed as an annual growth rate).

Table 1: Summary of Investment Flow Data

Industry	Equipment 1977	Equipment 1987	Computers 1977	Computers 1987
Total Manufacturing	82.2	83.6	7.4	14.1
Food products	78.1	80.3	3.8	4.8
Tobacco manufactures	68.7	80.6	6.7	11.6
Textile mill products	83.1	86.1	4.7	5.7
Apparel and textiles	75.9	78.3	8.7	10.9
Lumber products, ex furniture	82.6	84.2	2.2	3.7
Furniture and fixtures	68.2	69.5	7.2	12.3
Paper and allied products	90.1	87.1	7.1	10.9
Printing and publishing	81.7	82.7	14.9	17.1
Chemicals and allied products	84.7	82.5	7.5	13.1
Petroleum and coal products	73.1	76.2	8.0	14.8
Rubber and misc plastic products	81.5	85.4	3.4	5.5
Leather and leather products	77.4	77.7	7.7	11.1
Stone, clay, and glass products	82.0	88.7	5.3	9.8
Primary metal industries	85.8	88.9	5.1	11.2
Fabricated metal products	81.5	85.2	4.9	9.3
Machinery, except electrical	78.3	85.0	15.2	31.6
Electrical and electronic equip	82.6	86.5	11.1	19.4
Transportation equip	85.1	81.4	5.6	14.2
Instruments and related products	75.9	82.1	11.5	22.2
Misc manufacturing industries	76.3	82.8	8.3	14.4

Note: Equipment Share of Total Investment
and Computer Share of Equipment Investment

Table 2: Summary of Average Annual Capital Stock Growth

		1972-77	1977-82	1982-87	1987-91
Total Manufacturing	NEW	3.185	2.857	0.977	1.115
	OLD	3.408	1.951	1.838	.
Food products	NEW	2.586	1.896	0.694	0.770
	OLD	2.924	0.812	1.000	.
Tobacco manufactures	NEW	4.843	7.574	5.655	-0.426
	OLD	5.171	4.562	7.390	.
Textile mill products	NEW	1.894	-0.464	-1.127	-0.687
	OLD	3.055	-1.567	-1.273	.
Apparel and textiles	NEW	3.083	0.852	0.093	-0.402
	OLD	4.283	-0.172	0.245	.
Lumber products, ex furniture	NEW	4.657	1.256	-2.453	-2.049
	OLD	4.500	0.560	-2.486	.
Furniture and fixtures	NEW	3.823	2.387	2.132	1.654
	OLD	4.484	1.402	2.641	.
Paper and allied products	NEW	3.616	3.396	1.655	3.886
	OLD	3.250	2.208	1.905	.
Printing and publishing	NEW	2.119	3.458	4.204	3.714
	OLD	1.993	1.920	5.153	.
Chemicals and allied products	NEW	4.569	2.567	-0.776	2.003
	OLD	4.372	2.373	0.309	.
Petroleum and coal products	NEW	3.691	3.513	-0.002	0.361
	OLD	2.898	1.902	1.031	.
Rubber and misc plastic products	NEW	4.405	2.235	1.468	2.356
	OLD	5.887	0.542	1.126	.
Leather and leather products	NEW	0.478	1.070	-2.310	-2.668
	OLD	0.797	-0.641	-2.843	.
Stone, clay, and glass products	NEW	2.263	1.636	-1.310	-1.387
	OLD	2.597	0.835	-0.434	.
Primary metal industries	NEW	1.654	0.570	-1.941	-1.552
	OLD	1.377	-1.150	-3.391	.
Fabricated metal products	NEW	3.445	3.022	1.139	0.224
	OLD	3.398	1.473	0.619	.
Machinery, except electrical	NEW	4.664	5.580	2.071	1.036
	OLD	4.728	4.940	4.910	.
Electrical and electronic equip	NEW	3.610	5.966	5.840	3.320
	OLD	4.279	4.261	7.620	.
Transportation equip	NEW	2.162	3.755	2.096	0.567
	OLD	3.306	4.538	3.542	.
Instruments and related products	NEW	4.972	5.264	3.791	3.325
	OLD	5.484	4.153	5.954	.
Misc manufacturing industries	NEW	3.455	1.873	0.142	0.165
	OLD	5.070	1.450	0.979	.

Table 3: Summary of Average Annual Material Price Changes

		1972-77	1977-82	1982-87	1987-91
Total Manufacturing	NEW	10.151	9.222	-0.367	2.817
	OLD	9.820	9.698	0.370	.
Food products	NEW	7.993	6.861	0.582	2.948
	OLD	7.606	7.111	0.698	.
Tobacco manufactures	NEW	7.678	8.642	0.396	3.547
	OLD	7.493	8.264	0.703	.
Textile mill products	NEW	6.406	6.552	0.692	2.607
	OLD	6.400	7.285	0.752	.
Apparel and textiles	NEW	6.330	5.924	1.583	2.476
	OLD	6.645	6.129	1.985	.
Lumber products, ex furniture	NEW	10.302	4.936	2.165	3.457
	OLD	9.804	5.623	2.556	.
Furniture and fixtures	NEW	9.407	6.807	1.613	2.694
	OLD	8.923	7.185	2.312	.
Paper and allied products	NEW	11.434	8.197	1.467	2.916
	OLD	10.510	7.976	1.956	.
Printing and publishing	NEW	9.524	7.848	2.092	2.573
	OLD	8.295	7.938	2.768	.
Chemicals and allied products	NEW	13.076	9.074	-0.474	4.249
	OLD	11.749	9.642	0.809	.
Petroleum and coal products	NEW	16.808	18.030	-9.412	3.227
	OLD	19.356	21.003	-7.095	.
Rubber and misc plastic products	NEW	11.543	8.473	0.826	3.207
	OLD	10.251	8.483	1.656	.
Leather and leather products	NEW	8.373	7.940	2.580	3.662
	OLD	7.630	8.191	3.276	.
Stone, clay, and glass products	NEW	11.273	10.024	0.815	2.630
	OLD	9.828	9.432	1.421	.
Primary metal industries	NEW	11.809	8.287	0.740	2.825
	OLD	11.039	7.825	0.992	.
Fabricated metal products	NEW	10.859	8.069	1.139	2.636
	OLD	10.084	8.073	1.722	.
Machinery, except electrical	NEW	9.328	7.539	0.260	2.178
	OLD	8.432	7.412	0.356	.
Electrical and electronic equip	NEW	7.989	7.463	1.124	1.959
	OLD	7.732	7.494	2.148	.
Transportation equip	NEW	9.381	8.748	1.293	2.492
	OLD	8.994	8.815	1.589	.
Instruments and related products	NEW	9.039	7.287	1.225	2.328
	OLD	8.108	7.638	2.084	.
Misc manufacturing industries	NEW	9.968	7.959	1.579	2.529
	OLD	9.123	8.066	2.053	.

Table 4: Summary of Average Annual Price Changes

	1972-77	1977-82	1982-87	1987-91
BEA PDE Deflators				
Total Computers	-15.233	-13.838	-17.224	-10.722
Mainframe	-9.265	-32.369	-19.416	-9.132
PC & Workstation	.	-32.368	-19.564	-12.729
Disks etc.	-20.030	-15.521	-10.292	-8.704
Printers	-6.156	-7.072	-17.518	-6.993
Displays	-3.764	-8.791	-22.198	-15.128
Prices for 3573				
BW : Benchmark Years	-14.294	-13.833	-17.263	-8.539
Hed : Hedonics	-14.294	-18.534	-17.777	-9.978
NoH : No Hedonics	-0.280	-4.284	0.060	-2.026
F87 : Fixed 1987	-14.294	-8.343	-15.896	-10.247
Gray: Fixed 72	-14.294	-13.833	-18.629	.
Hedonic 4-Dig				
SIC 3579	4.973	-7.038	-8.095	-2.990
SIC 3661	9.678	-3.799	-6.989	-6.276
SIC 3674	-0.280	-26.711	-17.477	-8.139
Total Mfg., Divisia Weighted Agg				
BW :Benchmark	8.934	8.000	0.510	2.783
Hed : Hedonics	8.934	7.823	0.342	2.592
NoH : No Hedonics	9.056	8.124	0.890	2.925
F87 : Fixed 87	8.975	8.133	0.602	2.672
Gray: Fixed 72	8.975	8.052	0.418	.
BW 4-digits				
Fixed 87 Agg	3.044	5.826	0.276	2.867
Implicit Agg	9.193	8.354	0.654	2.717

Table 5: Summary of Average Annual TFP Growth

	1972-77	1977-82	1982-87	1987-91
Domar Weighted				
BW : Benchmark Years	0.355	-0.374	1.882	-0.511
Hed : Hedonics	0.294	-0.224	2.009	-0.285
NoH : No Hedonics	0.146	-0.583	1.490	-0.623
F87 : Fixed 1987	0.275	-0.575	1.763	-0.371
Gray: Fixed 1982	-0.144	0.255	2.215	.
Value Added, Divisia				
BW : Benchmark Years	0.420	-0.336	1.219	-0.572
Hed : Hedonics	0.381	-0.123	1.391	-0.357
NoH : No Hedonics	0.249	-0.518	0.846	-0.669
F87 : Fixed 1987	0.376	-0.470	1.116	-0.448
Gray: Fixed 1982	-0.069	-0.044	1.291	.
Value Added, Fixed 1987				
BW : Benchmark Years	0.295	-0.376	1.189	-0.473
Hed : Hedonics	0.231	-0.386	1.293	-0.214
NoH : No Hedonics	0.308	-0.352	0.916	-0.618
F87 : Fixed 1987	0.287	-0.399	1.112	-0.334
Gray: Fixed 1982	-0.210	-0.060	1.223	.

Table 6: TFP Growth Rates, 1977-1988, Domar Weighted

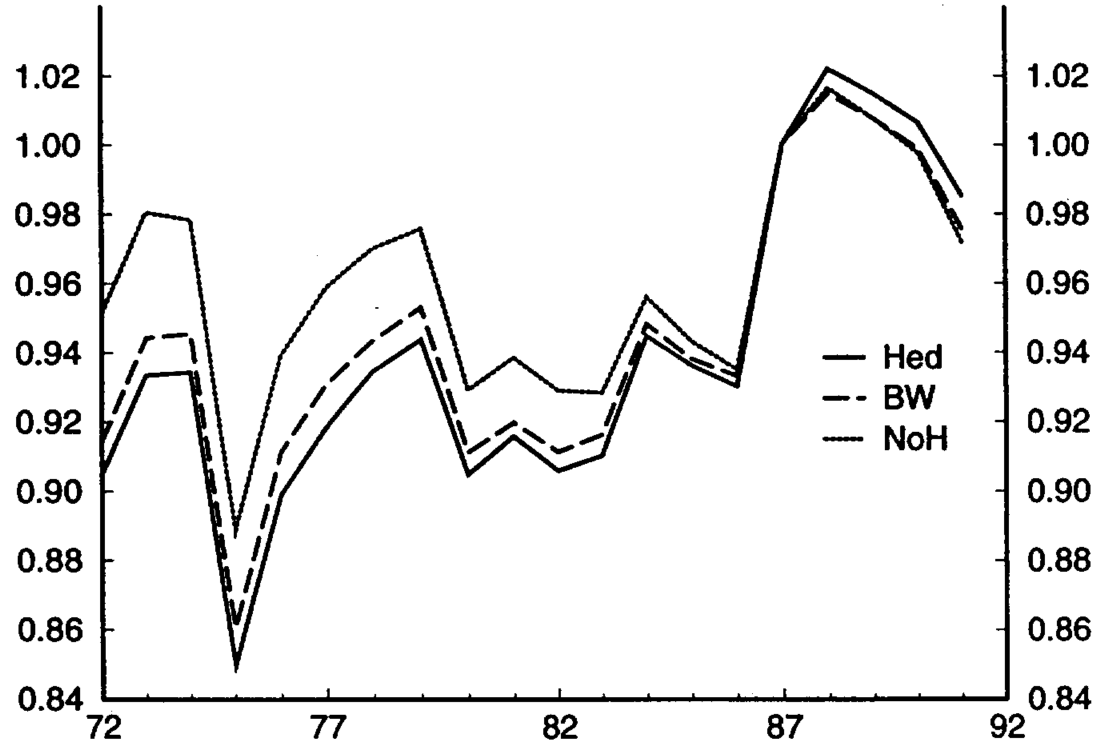
Industry	BW	Hed.	NoH.	F87	Gray
Total Manufacturing	0.754	0.893	0.453	0.594	1.235
Food products	0.552	0.552	0.542	0.552	0.662
Tobacco manufactures	-0.435	-0.430	-0.439	-0.430	-0.217
Textile mill products	0.094	0.054	0.069	0.054	0.238
Apparel and textiles	0.057	0.036	0.031	0.036	0.129
Lumber products, ex furniture	-0.089	-0.253	-0.128	-0.253	0.040
Furniture and fixtures	-0.352	-0.338	-0.389	-0.338	-0.289
Paper and allied products	-0.042	-0.039	-0.052	-0.039	-0.005
Printing and publishing	-0.035	-0.051	-0.049	-0.051	0.065
Chemicals and allied products	-0.231	-0.278	-0.281	-0.278	-0.162
Petroleum and coal products	0.027	-0.032	0.005	-0.032	0.836
Rubber and misc plastic products	0.138	0.194	0.125	0.195	0.240
Leather and leather products	-0.356	-0.344	-0.366	-0.344	-0.218
Stone, clay, and glass products	-0.427	-0.418	-0.435	-0.418	-0.432
Primary metal industries	-0.296	-0.300	-0.302	-0.299	-0.296
Fabricated metal products	-0.429	-0.429	-0.427	-0.423	-0.328
Machinery, except electrical	0.226	0.356	-0.225	0.021	0.209
Electrical and electronic equip	0.156	0.675	0.164	0.112	0.281
Transportation equip	-0.484	-0.498	-0.466	-0.479	-0.491
Instruments and related products	0.236	-0.136	0.228	-0.017	0.207
Misc manufacturing industries	-0.493	-0.553	-0.472	-0.550	-0.513

Note: BW : Benchmark years weighted
Hed.: Added Hedonics
NoH.: No Hedonics
F87 : Fixed 1987 weights
Gray: Old Dataset (82 Fixed)

Figure 1

Total Factor Productivity Measures

Domar Weights: Total Manufacturing



Percentage Change in TFP

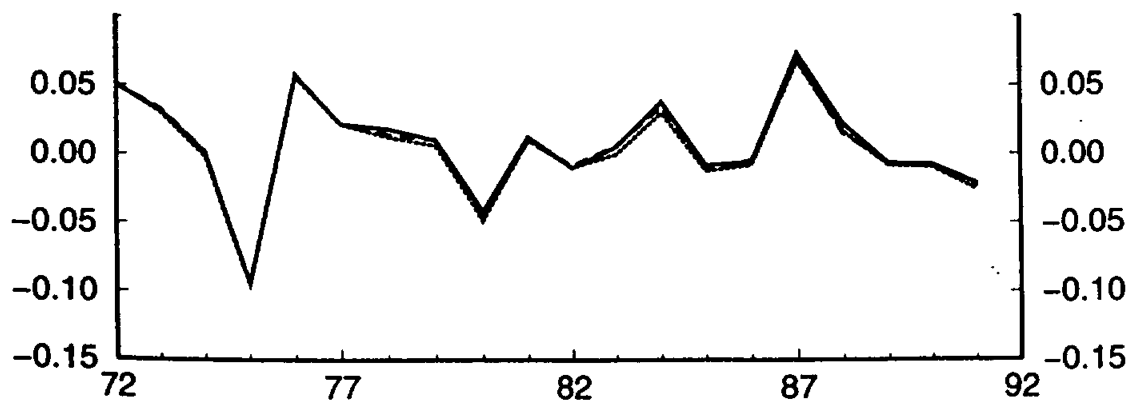


Figure 2

Total Factor Productivity Measures