

**Finance and Economics Discussion Series  
Divisions of Research & Statistics and Monetary Affairs  
Federal Reserve Board, Washington, D.C.**

**A Consistent Accounting of U.S. Productivity Growth**

**Eric J. Bartelsman and J. Joseph Beaulieu**

**2004-55**

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

## **A Consistent Accounting of U.S. Productivity Growth\***

Eric J. Bartelsman

ebartelsman@econ.vu.nl

*Free University of Amsterdam and Tinbergen Institute*

and

J. Joseph Beaulieu (contact)

joe.beaulieu@frb.gov

*Board of Governors of the Federal Reserve System*

Industrial Output Section, Division of Research and Statistics

Washington, DC 20551

September 2004

### **Abstract**

This paper is an exploration in the relative performance and sources of productivity growth of U.S. private businesses across industries and legal structure. In order to assemble the disparate data from various sources to develop a coherent productivity database, we developed a general system to manage data. The paper describes this system and then applies it by building such a database.

The paper presents updated estimates of gross output, intermediate input use, and value added using the BEA's GPO data set. It supplements these data with estimates of missing data on intermediate input use and prices for the 1977-1986 period, and it concords these data, which are organized on a 1972 SIC basis, to the 1987 SIC in order to have consistent time series covering the last twenty-four years. It further refines these data by disaggregating them by legal form of organization. The paper also presents estimates of labor hours, labor quality, investment, capital services and, consequently, multifactor productivity disaggregated by industry and legal form of organization, and it analyzes the contribution of various industries and business organizations to aggregate productivity. The paper also reconsiders these estimates in light of the surge in spending in advance of the century-date change.

JEL Codes: D24, E23

Keywords: databases, legal form of organization, productivity

\* The authors would like to thank Carol Corrado for comments at various stages of the project and Jonathan Eller, Suzanne Polatz, Marcin Przybyla, Brian Rowe, Koen Vermelyen, and Matt Wilson for research assistance, the BEA for providing data, and comments from Barbara Fraumeni, Edward Prescott, and Daniele Coen-Pirani, and participants at the NBER/CRIW Summer Institute, Federal Reserve Bank of Minneapolis, and the Winter Meetings of the Econometric Society. This paper represents the authors own views and not those of the Board of Governors of the Federal Reserve System or its staff.

## I. Introduction

Zvi Griliches thought so much of the inherent difficulty in building data sets suitable for analysis that he devoted a chapter to the problem in the *Handbook of Econometrics* (1986). With respect to available data, he wrote:

There are at least three interrelated and overlapping causes of our difficulties: (1) the theory (model) is incomplete or incorrect; (2) the units are wrong, either at too high a level of aggregation or with no way of allowing for the heterogeneity of responses; and, (3) the data are inaccurate on their own terms, incorrect relative to what they purport to measure. The average applied study has to struggle with all three possibilities (pp. 1468-69).

The problems are especially acute in the study of productivity, where researchers usually have to "find" their data from different sources. These disparate data sources are produced by different government agencies or private outfits and are designed to answer different questions. As such, changes in the ratio of real outputs to inputs may reflect inconsistencies in the dataset, rather than movements in productivity. Because the source data were not assembled to study productivity, they can also be incomplete.

A basic measure of TFP requires data on deflated output, labor input, capital services, intermediate inputs, and the distribution of income to factors of production. These data need to measure the activities of the same producers to get an accurate reading of productivity growth but, if they are assembled from different sources, they may not be comparable for a variety of reasons, even if the descriptions of the activities covered are the same. Across different variables, the underlying micro data may have been collected in an inconsistent manner. Data sets may be organized using disparate classification systems. Even if the data come from one source, the classification system can vary over time. Sometimes different classification schemes reflect more fundamental differences, but the data still may contain exploitable information.

Another problem is the estimation of missing data. Data are often published at higher levels of aggregation than is desired. Sometimes data are available at a fine level of detail in one dimension but only at a very high level of aggregation in another dimension when detailed data are needed in both dimensions. A practical problem occurs when new data releases provide totals first, only to follow with detailed datasets at a considerable lag. In order to conduct research with all the desired detail and to make use of the latest data, procedures are needed to best use all the available information.

The purpose of this paper is threefold. First, it describes a system that we have built in order to overcome the data hurdles just described. We have developed a general approach to data organization and a set of standardized routines that allow us to produce consistent datasets from multiple sources. In part, the system allows researchers to create coherent data out of various bits of information. Second, the paper sketches the use of the system to construct a dataset to study productivity. Third, the paper presents some simple applications. It reports estimates of TFP growth by industry and by legal form of organization, and it reconsiders these estimates assuming that firms scrapped an unusual amount of capital addressing potential Y2K bugs.

Productivity in the United States is the focus of the paper and is prominent in the discussion of the data problems. However, the data issues are ubiquitous, and the system can accommodate other areas of study, such as international trade, economic geography, macroeconomic income dynamics, and supply and demand in general equilibrium. The paper emphasizes industry-level data in its analysis, but the system easily scales down to more micro data or scales up to more aggregate data, such as cross-country data.

The advantages of such a system are numerous. Obviously, it allows economists to

entertain daunting research projects. Moreover, documentation of any particular application can easily be thorough because the use of standardized routines permits a terse description of complex operations. Also, because of the automated nature of the system, researchers can easily vary the particular assumptions that they used to create their estimates in order to test for their sensitivity. A researcher could go farther and produce confidence bands around such estimates. One could also apply this methodology to already published data that are produced by the statistical agencies. Indeed, some datasets that have been made available by government agencies rely on the same techniques available in our system to produce their estimates. As such, a reconsideration of the assumptions that these agencies employ to produce their estimates may be useful.<sup>1</sup> Finally, this approach allows one to consider rigorously counterfactual exercises or to explore the implications of mismeasurement, such as in Jorgenson and Stiroh (2000).

## **II. Description of the System**

The system that we have developed provides a practical method to cope with the data problems. Before describing our approach, we give some brief examples of the types of hurdles faced when building a consistent dataset.

### *II.A. Data hurdles*

To start, a potentially difficult problem arises when two aggregates that are titled the same in publications are defined differently. For example, before the 2003 comprehensive benchmark

---

<sup>1</sup> With the exception of the recent literature considering mismeasured or biased prices, we are not aware of a lot of papers that directly explore the idea that published data are partially built on assumptions and models where alternatives can be considered. Exceptions include Wilcox (1992) and Miron and Zeldes (1989). There is, however, a developed literature studying the effects of measurement error; see Bell and Wilcox (1993) and references therein. See Weale (1985) for an approach similar to the strategy that could be contemplated with our system.

revision to the NIPAs, the BLS and the BEA had different definitions of nonfarm business. The BLS (and now the BEA) excludes two imputations (owner-occupied housing and the rental value of non-profits' capital equipment and structures) that the BEA makes to estimate GDP.

Although a careful reading of underlying documentation can trap such differences, only the detailed reconstruction and re-aggregation of the underlying data will allow one to reconcile the differences in outcomes of analysis based on the two output definitions.

A more fundamental problem is related to the underlying data collection. One well-known example of this is the firm-establishment problem (Postner, 1984). U.S. business data are usually collected at one of two levels: at the establishment level, such as at individual plants, stores, or comparable places of work, or at the firm level. A problem arises, however, when a firm has multiple establishments that are engaged in different lines of work. GE has extensive operations in manufacturing, finance, and services. Data collected at the establishment level, will effectively split GE data among different industries along the different activities of the individual establishments. Data collected at the firm level will classify all of GE in one industry based on its major activity. Currently Compustat assigns GE to the catchall category miscellaneous, although a few years ago GE was designated an electrical equipment manufacturer.

Researchers manipulating the data have to know how the data were collected. In putting together the Gross Product Originating (GPO) dataset, economists in the Industry Division at the Bureau of Economic Analysis (BEA) have converted all of the data to an establishment-basis concept.<sup>2</sup> The NIPAs, on the other hand, also present some industry data, but they are not

---

<sup>2</sup> In this paper we refer to these industry data as the GPO dataset. The BEA has published versions of this dataset under different names. In 2003, the dataset was called the GDP-by-Industry data. In 2004, the

consistent across different types of income. The NIPA compensation data are collected at the establishment level, and as table 1 illustrates, the two sources match. The NIPA profit data, however, are collected at the firm level from administrative sources, and therefore, the two databases do not agree on the mix of profits across industries, although they do match in the aggregate.

A problem that is particularly annoying to researchers is when two different classification schemes are used. Researchers often want long time series, but classification schemes evolve over time. Usually, industry data before 1987 are based on the 1972 Standard Industrial Classification System (SIC), while data afterwards are organized on the 1987 SIC. Recently statistical agencies have begun to switch to the North American Industry Classification System (NAICS). Input-output tables use their own classification systems, which also have changed over time. Reclassifying the data so that they are all on one system — a procedure called *concording* — can be difficult when only published data sources are available.

Sometimes, two classification systems may be motivated by entirely different concepts. Nevertheless, incorporating information from both systems may be useful. A good deal of NIPA data is presented by legal form of organization (LFO). While these data cannot be simply linked to data split by industry, we know from the economic censuses that the mix of corporate versus noncorporate businesses varies across industries. Manufacturing, mining, and utilities are predominately corporate, while some service industries, such as membership organizations, personal services, and legal services have a large fraction of unincorporated firms. As such, the LFO data contain exploitable information on the mix of an aggregate across industries.

---

BEA significantly changed the method by which it calculates these data, yet it continues to refer to them as the GDP-by-Industry data. In order to clearly identify the data that we are using as the data consistent

Another way data can be mismatched to the needs of the researcher is when some data are incomplete or missing. One dataset may present manufacturing industry data split at the two-digit SIC level, while another may include only durable and nondurable sub-aggregates. A different example can be found in the NIPAs where at the national level, indirect business taxes, business transfers, and the surplus of government enterprises less subsidies are presented separately, but for corporations only the sum is listed.

A second example of the problems presented by missing data arises when a researcher has data of aggregates in different dimensions but does not have detailed estimates broken out in each dimension. For instance, the GPO contains information on noncorporate net interest paid by various industries. The NIPAs provide national totals for net interest paid by partnerships and proprietorships and by other private businesses. No published data, however, exist on net interest paid split both by industry and by this level of legal form of organization.

A final way in which data can be incomplete is when aggregate data are updated, but updated disaggregated data are not yet available. For example, the BEA publishes initial data on all of the components of gross domestic income for a particular year at the end of March of the following year. Typically, it publishes benchmarked data at the end of July, but the industry data are not finalized until well after. One could imagine that it would be possible to develop initial estimates for the recently completed year and incorporate the revised national data to update quickly industry estimates in prior years. Indeed, the BEA has developed a program to produce such “accelerated current-dollar estimates” (see Yuskavage, 2002); even so, revised data at the more detailed level are only available with the release of the full dataset.

---

with the 2003 and prior methods, we use the old nata. GPO data.



## *II.B. Overview of the Data System*

The system that we have developed can be thought of as comprising four, interrelated components that provide practical tools to deal with these problems. First, we re-code and store economic data, such as the NIPAs, GPO, and input-output data in a relational database. Relational databases are structured differently than hierarchical databases, which are commonly used by economists either for research or for sourcing time series from commercial data vendors. Hierarchical databases were thought to be particularly well suited for time series built up from units related in a hierarchical manner. The mnemonics given to the series for retrieval from the database reflect the hierarchy. In the early years of relational databases, these structures were thought to be difficult to manipulate efficiently (Codd, 1990). However, the hierarchical structure can be stored in related tables in the form of an ordered graph and can be traversed with recursion and manipulated using queries.

This brings us to the second feature of our system. The ways in which the data inter-relate are coded in meta databases. In particular, these meta data describe how the detailed data aggregate in a hierarchy and how two classifications map into each other through concordances. The meta data form linear restrictions across observations that ensure overall consistency of the dataset.

The relational database and the meta data make it possible to write standardized routines, or tools, to manipulate the data. We think of these generalized tools as falling in one of four categories: aggregating, disaggregating, balancing, and concurring data. These four operations help to overcome many of the hurdles that researchers face when using data from different sources.

Finally, the system contains some specialized tools necessary for the study of

productivity. These specialized tools allow users to estimate capital stocks, capital services, and total factor productivity employing a variety of assumptions.

### *II.C. Relational databases*

A relational database is an organization of data that takes seriously the idea that a piece of datum is not simply the particular numerical value an observation takes but all of the characteristics that identifies the observation. For example, suppose we have these pieces of data:

the logging industry purchased \$5.2 billion of forestry products in 1992;  
 or  
 non-profit health care organizations employed 5.2 million workers in 1995.

Besides the number 5.2, six other characteristics, or *dimensions*, describe that particular observation. In a relational database, these observations may be coded as:

<b>Relational Database</b>						
<b>Sector</b>	<b>Industry</b>	<b>Product</b>	<b>Transaction</b>	<b>Date</b>	<b>Units</b>	<b>Value</b>
private	logging	forestry	intmd. input	1992	bil. dollars	5.2
non-profit	health	workers	labor input	1995	mil. employees	5.2

where each dimension gets its own column to contain data.

This is in contrast to the more typical time-series database, which may look something like:

<b>Inputs Used by Industry</b>		
<b>Date</b>	<b>p241_03001</b>	<b>n80_emp</b>
1992	5.2	:
1995	:	5.2

In the second type of database, a lot of information is contained in the mnemonics for the variables: ‘p’ indicates private business, ‘241’ is the (SIC) code for the logging industry, and ‘03001’ is the (IO) code for forestry products. Other information is included in the name of the

database, in this case that all of the information in the dataset pertains to input usage. Other information may not be noted at all; somehow the user knows that the units of employees are millions of workers, while intermediate goods are measured in billions of dollars.<sup>3</sup> In some systems, this information is contained in attached labels, attributes, or in separate documentation.

A relational database has a couple of advantages. One can search over the particular values of a particular column and operate on a particular subset of values. For instance, for each value of the triplet (sector, date, units), one can sum over the values where activity = “logging” and transaction = “intmd. input” to get total intermediate input usage by the logging industry in each year by legal form of organization. To associate two databases that are organized the same way, one can code the identifying characteristics in the same manner and then join the datasets. The ubiquitous language that is used to make such calculations is the Structured Query Language (SQL); various database packages have their own particular implementations.

The example of the relational database presented above contains all of the dimensions that we use in our application to the study of productivity, except one additional dimension. In the NIPAs, two types of data are largely imputed. First, a large component of consumption is owner-occupied housing. The BEA accounts for this by assuming that there is an entity that owns the stock of owner-occupied housing and rents it back to its owners. The rental value of owner-occupied housing is treated as consumption. To preserve the identity that Gross Domestic Product equals Gross Domestic Income (up to the statistical discrepancy), the BEA imputes income to this entity. The second imputation involves the rental value of non-profit’s capital equipment and structures. As for owner-occupied housing, the BEA pretends that there is an

---

<sup>3</sup> While it may seem trivial, most databases do not store the ‘full’ number (*e.g.* 5,200,000) but instead truncate the number at a standard level. However, this level is not standard across datasets, and

entity that owns this capital and rents it to non-profit organizations.<sup>4</sup> In the case of housing, this technique also makes GDP invariant to whether the house is rented or occupied by its owners. We refer to this additional dimension as **Imputed**, and let it take one of three values: “owner-occupied housing”, “rental-value of non-profits”, or “not imputed”.

The dimensions that we use in this study are summarized below:

<b>Sector</b>	represents the NIPA institutional sectors (business, general government, households, and non-profit institutions). The business sector is further refined by legal form of organization (corporate, noncorporate, etc.).
<b>Industry</b>	describes the particular industry that defines the producer, such as agriculture, manufacturing, etc. Sector and industry are not the same concept, but they are sometimes related. There are numerous classification schemes for industries, such as NAICS, NACE, ISIC, SIC, etc.
<b>Imputed</b>	accounts for whether the data apply to the two imputed sectors, owner-occupied housing and the rental value of non-profits’ capital equipment and structures, or not.
<b>Product</b>	represents the type of goods or services produced, purchased, consumed or supplied. As with industry there are multiple classification schemes, such as input-output commodities, 5-digit products in the 1987 SIC, and the types of purchased capital goods in the capital-flows survey, etc.
<b>Transaction</b>	describes where the product or input relates in the chain of production. There are two types of transactions, distributive and productive. Distributive transactions are typically income, or income-like items such as compensation, profits, indirect business taxes, dividends paid, capital consumption, etc. Productive transactions relate to goods and services produced or consumed as inputs to the productive process, such as gross output, intermediate inputs, labor hours, capital services, investment, consumption, etc.
<b>Date</b>	is the particular date of observation, such as 1993, 01:Q1, January-1994. Note that the date dimension also can be coded to incorporate information on the frequency of the data (monthly, quarterly, etc.) and other timing attributes,

---

sometimes is not consistent within a dataset.

<sup>4</sup> To be exact, these are only non-profit institutions that primarily serve individuals. Non-profits that primarily serve businesses, such as trade associations, are treated like any other business in the NIPAs in that their consumption of nondurable goods are counted as intermediate usage and their purchases of equipment and structures are counted as investment. The income paid by these institutions to various factors of production is included in the aggregates for corporations.

such as average over period, end of period, etc. One could imagine that in some applications, these attributes would be coded in different dimensions when the dataset contains a variety of such attributes.

- Unit** describes how the variable is measured and whether it is a nominal variable, price deflator or real variable. Examples include millions of dollars, Fisher chain-weighted price index, 1996=100, etc.
- Value** reports the numerical value of the data of interest.

A complete accounting of the circular flow of goods, services, and income would include a few other dimensions that identify not only who produces the good or service or who pays the income, but also who purchases the good or service or receives the income. In such a way, one could fully integrate all of the NIPA data into the system (for example, tables of personal income and government receipts and expenditures). Such an analysis would be necessary when studying income dynamics or general equilibrium.

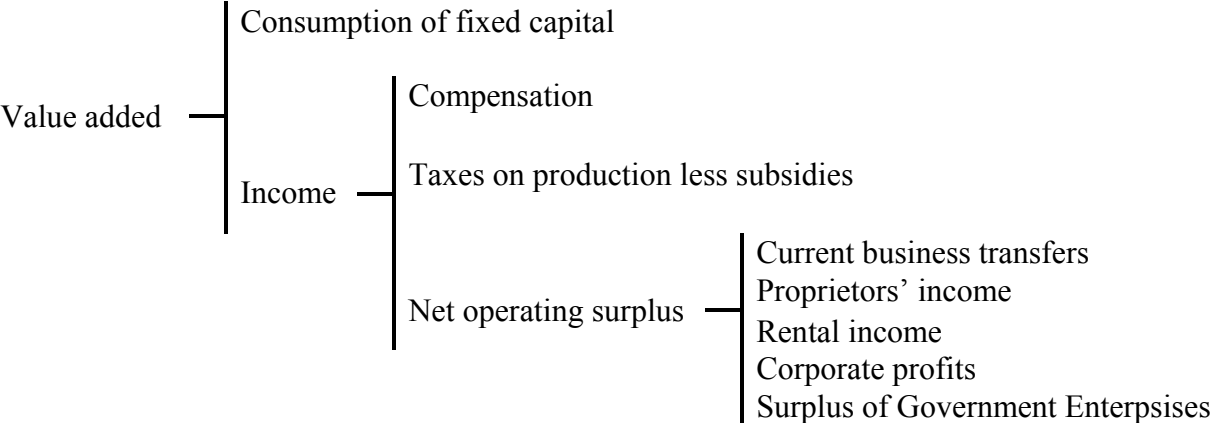
The presence of various industrial classification schemes presents a small dilemma. One could imagine having separate dimensions to describe each classification scheme: one for SIC 1972, another for SIC 1987, and a third for NAICS. Under this strategy, observations using one system would be concorded to all of the other relevant classification systems before they were stored in a database. We do not follow this strategy. Usually one is not particularly interested in seeing how the different classification systems compare; instead, one just wants to convert all of the data to one particular system. Maintaining new data on an old classification scheme could become burdensome, and the newer system should have some advantages in representing the current structure of the economy. Nonetheless, it would be possible to implement this strategy, and in some cases, such as building a concordance from micro data, it would be the way to go.

#### *II.D. Meta data*

The system requires two types of meta data, hierarchies and concordances. A hierarchy

describes how the data add to their total. Knowing the hierarchy is useful for several reasons. It makes possible the calculation of interesting subaggregates, and it makes matching datasets that differ on their level of aggregation easier. One can keep some subtotals in a database and use the hierarchy to then exclude those subaggregates when calculating other subaggregates or totals. It may be important to carry these subaggregates in the database, especially when they are read directly from a data source. Rounding issues make the subaggregates read directly from the data source more accurate than anything that can be calculated, especially for chain-weighted aggregates.

Finally, and perhaps most importantly, the hierarchies code the myriad of linear constraints that exist in economic theory, as well as various datasets. For instance, the fact that value added is the sum of various income components can be represented by the following transactions hierarchy:



Suppose at the same time, there is another hierarchy that divides GDP between the value added of manufacturing and nonmanufacturing industries. Then, the sum of the income components of manufacturing sum to value added of manufacturing, and likewise for nonmanufacturing. Manufacturing and nonmanufacturing compensation sum to total compensation, as well as for other types of income, and so on.

In some instances, one may need multiple hierarchies to represent ostensibly the same classification system. For instance, datasets sometimes organize government data by first splitting the data between the federal (FED) and state and local (SL) governments, and then by splitting each between general government (GG) and government enterprises (GE). Other times the order is reversed.

<b>HIERARCHY 1</b>				<b>HIERARCHY 2</b>			
GOV				GOV			
FED		SL		GG		GE	
GGF	GEF	GGSL	GESL	GGF	GGSL	GEF	GESL

Note that the lowest nodes, or *atoms*, of the two hierarchies are the same, and so a concordance between these two ways of organizing the data is a simple one-to-one match. Which hierarchy to use depends on which subaggregates are relevant. Another example of the rearrangement of the same atoms also involves the treatment of government enterprises. Sometimes, government enterprises are counted as a business; other times they are treated as part of the government.

It is obvious from the above examples that the construction of hierarchies and the number of dimensions are not unique. For example, as was suggested by the two hierarchies that describe the government, instead of having two hierarchies, one could create separate dimensions:

<b>METHOD 1</b>	<b>METHOD 2</b>	
<b>SECTOR</b>	<b>SECTOR 1</b>	<b>SECTOR 2</b>
GGF	GEN	FED
GGSL	GEN	SL
GEF	GE	FED
GESL	GE	SL

The obvious problem with creating multiple dimensions in this example is that for sectors other

than the government, the second dimension is irrelevant. Nonetheless, how to organize the dimensions and the hierarchies is a matter of choice.

The second type of meta data, a *concordance*, describes how two classification schemes relate. The concordance can be as simple as a list of which components in one system map to the components of a second system and vice versa, or it can provide more detail on the relative magnitudes of how much of one component of one system is split among the components of the other system. What distinguishes a concordance with detailed information on relative magnitudes from simply a detailed dataset is that the information on magnitudes in a concordance is typically available for only one year. The concordance tool ensures that these relative magnitudes are applied across years, and the discussion of the concordance tool describes concordances in more detail.

### *II.E. Standardized operations*

The third component of the system uses the meta data along with the organization of the data in a relational database to automate regular dataset operations.

#### *II.E.1. Aggregating*

The most straightforward operation is aggregation. Nominal dollar and count data, such as hours and employment, are simply added up a defined hierarchy to calculate aggregates at various levels of detail. Other types of aggregation, such as Laspeyres, Paasche, Fisher ideal, or Divisia chained indexes involve more complex operations that require additional data on weights.

#### *II.E.2. Disaggregating*

A second operation that is often required is disaggregation, which is the inverse operation of aggregation. Given an aggregate, one can estimate the constituent pieces. For instance, in the GPO data before 1987, industries 36 (electrical machinery) and 38 (instruments) are aggregated



together; however, we would like to split these in two.

The difference between aggregation and disaggregation is that the former is a many-to-one operation. No other information besides the constituent pieces, and perhaps corresponding weights in the case of fancier aggregates, is required to calculate an aggregate. On the other hand, disaggregation is usually a one-to-many operation, and thus, one needs additional information to choose among the infinite possible ways to split a total.<sup>5</sup> We refer to this additional information as a ‘pattern’. The pattern data need not be consistent with the original data of interest. After all, if the pattern data were to aggregate to the target, one would already have consistent estimates of the pieces.

For simple data that add, the procedure scales up or down the pattern data to yield disaggregated pieces that sum to the known total. Let  $i$  index the components of an aggregate  $T$ . Deont the observed aggregate  $d^T$ , and suppose that there are pattern data on the pieces,  $p^i$ . Then the estimate of the disaggregated pieces is given by  $d^i = d^T p^i / \sum_i p^i$ . In the case of Fisher-ideal indexes, the procedure does this separately for Paasche and Laspyres indexes, which do add, and then takes the geometric average of the two.

The quality of the result depends on how well the initial pattern reflects the true distribution of the aggregate. Sometimes, one only has a few scraps that one hopes get the rough magnitudes right; in the limit, the fall back could be to simply split the aggregate evenly, *i.e.*  $p^i = 1$ . Other times, some market conditions or other reasonable assumptions may be used to justify a particular pattern. For instance, suppose one has an aggregate of labor hours that is to

---

<sup>5</sup> In cases when the aggregate and all but one component are known, the procedure is exact, and no pattern data are needed. This case arises when one wants to exclude one component from a large aggregate; typically, all of the data on both the aggregate and the piece to be excluded are already known.

be split among two industries. Using observed compensation as pattern data of the two industries implicitly assumes that the hourly compensation rates in the two industries are the same. Industries that are aggregated together tend to be close in the classification scheme, and hence, similar in structure, opening up the possibility that this strategy can be effectively employed in a number of cases.

The automated nature of the tool provides a couple of advantages. By varying the pattern data, such as by adding random noise, one can measure how sensitive the results are to the original pattern. Indeed, with a set of statistical assumptions, one could estimate standard errors around these estimates.

### *II.E.3. Balancing*

A third operation, balancing, allows one to estimate data subject to linear constraints in multiple dimensions. An example of a balancing problem shows up when trying to calculate capital services. To do this, one needs investment by type of equipment and by type of industry, while only data on economy-wide investment by type of equipment and total investment by industry may be available.

Several solutions to this problem have been proposed in the literature (Schneider and Zenios, 1990). We offer three. The first is directly applicable when, as in the above investment example, there are linear constraints in two dimensions. In this particular example, one can think of the unknowns as a matrix, where the rows represent different values in one dimension, and the columns represent different values in the second dimension. For instance, the rows can represent different industries, while the columns could represent different asset types.

		Asset types			Row controls
		T <sub>1</sub>	T <sub>2</sub>	...	Totals
Industries	I <sub>1</sub>	$a_{11}$	$a_{12}$	...	$\sum_{j=1}^J a_{1j}$
	I <sub>2</sub>	$a_{21}$	$a_{22}$	...	$\sum_{j=1}^J a_{2j}$
	⋮	⋮	⋮	⋮	
Column controls	Totals	$\sum_{i=1}^I a_{i1}$	$\sum_{i=1}^I a_{i2}$		$\sum_{i=1}^I \sum_{j=1}^J a_{ij}$

The constraints are represented as restrictions on the sums across the rows and columns.

Suppose one has an initial guess of the matrix,  $A_{k-1}$ , which is not consistent with the row and column controls. The first technique, the so-called RAS procedure, estimates  $A$  through the following algorithm. One pre-multiplies  $A_{k-1}$  by  $R_k$  so that  $R_k A_{k-1}$  satisfies the column controls. Then one post-multiplies  $R_k A_{k-1}$  by  $S_k$  so that  $R_k A_{k-1} S_k$  satisfies the row controls. Let  $A_k = R_k A_{k-1} S_k$ . Repeating the procedure leads to a series of matrices that, under certain conditions, converges, so that  $A = R A S$ , where  $A$  satisfies both row controls and column controls.<sup>6</sup> The limiting condition,  $A = R A S$ , also explains the moniker ‘RAS’ algorithm that

<sup>6</sup> Bacharach (1965) provides conditions for uniqueness and convergence. Schneider and Zenios (1990) credit Sinkhorn (1964) for an early result that if the entries of  $A$  are strictly positive then the RAS

has been attributed to Stone (Stone and Brown, 1962). The restriction implied by the procedure that the final matrix is a function of only a series of row and column-scaling factors is also known as the biproportional constraint, and this algorithm is also known as biproportional matrix balancing.

A different strategy is to stack the columns of the matrix into a vector and write  $a_i^0 = a_i + \varepsilon_i$  or  $a_i^0 = \varepsilon_i a_i$  where  $a^0$  is the vector of initial guesses of the true value  $a$  and the error term  $\varepsilon_i$  has a known distribution. Two commonly used distributions are the normal and log normal distributions. The advantage of this approach is that it can handle multiple dimensions and more general restrictions. We further generalize the problem by allowing the constraints also to be measured with error.

The unknown values are estimated via a maximum likelihood procedure:

$$\begin{aligned} & \min_{a_i, v_k} \sum_{i=1}^N \frac{1}{\sigma_i} (\log(a_i) - \log(a_i^0))^2 + \sum_{k=1}^K \frac{1}{\sigma_k} (v_k - v_k^0)^2 \\ \text{or} \quad & \min_{a_i, v_k} \sum_{i=1}^N \frac{1}{\sigma_i} (a_i - a_i^0)^2 + \sum_{k=1}^K \frac{1}{\sigma_k} (v_k - v_k^0)^2 \\ & \text{both } s.t. \text{ the } k \text{ linear constraints } \sum_{i=1}^N \phi_i^k a_i = v_k \\ & \text{where } a_i^0, v_k^0, \sigma_i, \text{ and } \sigma_k \text{ are known.} \end{aligned}$$

If  $\sigma_k = 0$ , the control is measured exactly, and  $\lambda_k$  replaces  $1/\sigma_k$  in the minimization problem where  $\lambda_k$  is now an unknown Lagrangian multiplier to be solved for. Stone, Champernowne, and Meade (1942) first proposed a least squares model. In their application, they weighted the observations according to how precise the estimates of the pattern were, but they assumed that

---

procedure will converge.

the controls were measured exactly.

Each method has its own advantages. The advantages of the RAS model is that it is easy to calculate, and under certain circumstances, the biproportional constraint has been given an economic interpretation. In the case of calculating an input-output matrix in year  $t$  based on a known input-output matrix in year  $t-1$ , Parikh (1979) interprets the two scaling factors,  $R$  and  $S$ , as

- a *substitution effect* that measures the extent to which the output of a sector substitutes or has been substituted by the output of the product of other sectors as an intermediate input;
- a *fabrication effect* that measures the extent to which the ratio of intermediate goods to total gross output has changed in a sector.

The benefit from the statistical approach is that it allows one to test a subset of restrictions using either a likelihood ratio test or a Wald test. Weale (1985) uses this insight to test the hypothesis that the U.S. current account was positive in 1982-83 instead of negative, as measured by the BEA.<sup>7</sup> Modeling the distribution of the errors as a normal distribution, perhaps with a standard deviation proportional to the observed values of  $a^0$ , also allows the procedure to choose negative values. In cases where several values are known to be zero, a solution to the problem may require a switch in the signs of the initial guess, and in such a case, the RAS procedure will not converge.<sup>8</sup>

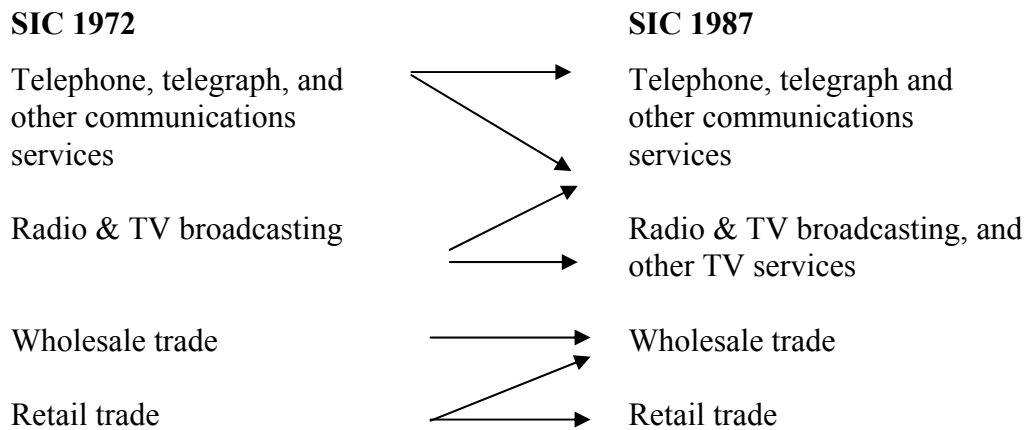
#### *II.E.4. Concoriding*

---

<sup>7</sup> Golan, Judge and Robinson (1994) develop a generalized version of the RAS model whereby the probabilities over a discretized space of values are estimated via something like the RAS procedure. These estimates also allow one to conduct statistical tests.

<sup>8</sup> The RAS procedure can be adapted to allow for negative values (Günlük-Şenesen and Bates, 1988) but the procedure will not switch the signs of the initial guesses. See Bartelsman and Beaulieu (2004) for a further discussion.

The last basic tool concords two data sets whose dimensions are organized on different classification schemes. For example, the GPO data from 1949 to 1987 are organized along the 1972 SIC; from 1987 to 2000 they are organized along the 1987 SIC. Some of these industries map to more than one industry, for example:



As is suggested by the above figure, the problem of concording data organized by the hierarchy on the left to the hierarchy on the right is simply to split the pieces of the left hand side into parts so that they can be allocated to the different categories on the right-hand side and then added back up. Concording the right-hand side to the left-hand side is the mirror image of this operation.

Thus, for the most part, the problem of concording is simply the organized use of aggregation and disaggregation. As such, the important part of the implementation is developing weights for the disaggregation. In most cases, information on the relative weights is limited because no data are reported on both bases, and thus, the weights have to be developed using whatever information is available. In concording the input-output tables to the GPO data, a few input-output industries had to be split; to do this, we used a variety of data, such as detailed

employment shares and census shipments data (see Appendix). To be sure, developing these weights can be difficult, yet is crucial to calculate an adequate concordance.

In one important case, data are reported on two bases, allowing for a richer concordance: the GPO data for 1987 are available using the 1972 SIC and the 1987 SIC. For example, industries 481,2,9 (telephone, telegraph, and other communications services) and 483-4 (radio and TV broadcasting) on the 1972 basis map to industries 481,2,9 (telephone, telegraph, and other communications services) and 483-4 (radio and TV broadcasting and other TV services) on the 1987 basis. One can think of the problem of developing these concordance weights as a balancing problem where the 1972 and 1987 totals are controls. As initial guesses for the pattern for all of the industries, we used the concordance in the NBER Productivity Database (Bartelsman and Gray, 1996) for manufacturing, and simply used 1/N for other industries for cells that are non-zero according to an available mapping. This simple example gave:

<b>Gross Output of Communications, 1987</b>			
		SIC 87	
		481,2,9	483-4
		157.8	42.1
SIC 72	481,2,9	170.1	157.8
	483-4	29.7	0.0
			12.3
			29.7

The cells of the matrix are the concordance weights. The advantage of balancing a matrix of weights is that one can concord data both ways in a consistent manner. Concoring data from the 1972 SIC to the 1987 SIC and then back again yields the original 1972 data.

Concoring provides a means for moving the data between two classification schemes in the same conceptual dimension. Technically analogous is the problem of cross-classification, such as moving data collected at the firm level and published by industry, to match data by

industry collected from establishments. The cross-classification table would contain data akin to that in a concordance, showing the amount in a firm-based industry that would split into various establishment-based industries.

### *II.F. Specialized Productivity Tools*

We have developed several tools to help in the study of productivity that we employ in the present study. One tool accumulates weighted levels of past investment using the so-called perpetual inventory method to estimate stocks of particular assets. The weights are modeled in the same manner as the BLS and FRB (Mohr and Gilbert, 1996) use to account for wear and tear, the average rate of discards, and the effects of obsolescence. A second tool weights these stocks using the standard user-cost model of Hall and Jorgenson (1967) in order to estimate capital services. The rate of return can be an ex-ante rate, such as a corporate bond rate, or an ex-post rate, such as property-type income divided by an estimate of the value of the capital stock. A third tool estimates TFP growth by calculating a Divisia index of the inputs using different approaches; the implementation in this paper uses cost shares to weight the inputs.

## **III. Creating a Data Set for the Study of Productivity**

### *III.A. Basic Industry Data*

The main component of the database that we have put together is the GPO dataset published by the BEA. The GPO dataset includes annual data on price deflators, real and nominal measures of gross output, intermediate inputs, and value added roughly at the two-digit SIC level. The dataset also includes nominal income components of value added, such as capital consumption allowances, compensation, and capital income. The data are consistent with the income-side



measure of national product in the NIPAs; the sum across all industries totals gross domestic income.<sup>9</sup> Data on employment and all persons engaged in production are also included in the dataset. Complete data are available from 1987 to 2001, where industries are classified by the 1987 SIC. Most measures are also available from 1977 to 1987 on the 1972 SIC basis, although some pieces are missing. Nominal data series and employment data extend back to 1949.

We made two types of adjustments to the GPO data. First, we concorded the data on the 1972 basis to the 1987 SIC to get a consistent time series. Second, we filled in some missing data.

In order to fill out some of the price data for 1977 to 1987, we concorded the 1982, 1987, and 1992 input-output tables to the GPO data (see Appendix). We used the implicit weights in the input-output tables to calculate price deflators for intermediate inputs; along with available gross product deflators, these yielded gross output deflators.

The GPO dataset includes data on all persons engaged in production, which equals the number of employees in an industry plus the number of people working alone. The BLS publishes aggregate estimates of the labor hours of the self-employed and an estimate of self-employed compensation. This last measure represents the fraction of proprietors' income that could be considered labor compensation, as if the proprietor pays a salary to him or herself. The BLS makes this calculation in order to correctly weight the contribution of labor and capital in production function estimates. We make this same adjustment at a more detailed level; we estimate self-employed hours and compensation by industry controlled to the BLS's aggregates.

### *III.B. Legal Form and Other Refinements*

---

<sup>9</sup> Gross domestic income equals gross domestic output less the statistical discrepancy, see (Parker and

We have further refined these industry estimates by splitting output between businesses and non-profits, and we further split businesses between corporate and non-corporate organizations.<sup>10</sup>

Splitting the industry output by legal form is useful because it better matches the sources of at least some of the income components. Much of the income data are collected through tax records, and corporations and other businesses file different forms. The source data are also adjusted for misreporting; the dollar adjustment to proprietors' income was more than twice as large as to corporate profits in 1996, even though proprietors' income is a much smaller fraction of national income (Parker and Seskin, 1997). This suggests that the measurement of output for the noncorporate sector is subject to larger errors than for the corporate sector.

Corrado and Slifman (1999) showed that productivity in the noncorporate business sector was measured to have been declining for over two decades, even though capital income as a share of output was relatively high. They pointed to mismeasured prices as one likely explanation for the confluence of these observations. To the extent that prices are biased upwards in industries that have a disproportionate share of noncorporate business, the real output of noncorporate business would be biased down more than for corporate business. Splitting individual industries by legal form — where presumably the output and input prices to the sectors within an industry are similar — and comparing their relative performances may shed some additional light on the issue.

### *III.C. Investment and Capital Stocks*

The investment series that we use are the detailed industry estimates of industry investment by

---

Seskin, 1997).

<sup>10</sup> See Seskin and Parker (1998) for definitions of corporations, sole proprietorships and partnerships, and other forms of legal organization as used in the NIPAs.

asset type that the BEA has made available on its web site. We have refined these data by splitting industry investment between corporate and noncorporate investment for each type of equipment and structure, controlling the total for each legal form to equal the data available in the Standard Fixed Asset Tables and the residential investment tables of the Detailed Fixed Asset Tables. The nonresidential investment tables report investment in equipment and in structures by legal form, divided among three activity groups (farm, manufacturing, and other). To refine these data by industry and by asset type, we used total industry investment by industry and by asset type as an initial pattern in our balancing routine. A practical problem in working with the data was that the investment figures were rounded to integers. In early years, or for activity/type combinations with low levels of investment, dividing nominals by reals provided a poor estimate of the deflator. To rectify this, we assumed that these asset prices did not vary by activity and used the deflator calculated from aggregate data.

Capital stocks are calculated by accumulating the investment data using the standard BLS stochastic mean-service life and beta-decay parameters. We estimate capital services using the Hall-Jorgenson formula using ex-ante returns, and to analyze trends, we separate capital services into three categories, high-tech equipment and software (ICT), other equipment, and structures.<sup>11</sup>

---

<sup>11</sup> ICT capital is defined as computers and peripheral equipment, communications equipment, photocopy and related equipment, instruments, and software.

### III.D. Labor Services

Analogous to capital, a unit of labor of a certain type may provide a different level of service than a unit of labor of another type. The measure of labor input appropriate for productivity analysis, labor services, is computed as a quality-weighted aggregate of labor hours by type. The weights used to aggregate labor are expenditures shares for each type.

For each industry and sector, information is thus needed on hours worked and compensation for workers by type. These data are not directly available from firm- or establishment-based data sources. However, the Current Population Survey (CPS) March Supplement from the U.S. Bureau of the Census has information on wages of workers, along with other worker characteristics such as age, gender, occupation, education, and industry.

To calculate labor services, we first estimated Mincer's wage equation of the following form:

$$\log(w(s,x)) = \text{const} + \alpha * s + \beta_1 * x + \beta_2 * x^2 + \beta_3 * x^3 + \beta_4 * x^4 + Z\Gamma + \varepsilon ,$$

where  $w(s,x)$  represents wage earnings of someone with  $s$  years of schooling and  $x$  years of work experience. In the regression we also included gender, part-time/full-time and ICT occupation dummies, summarized in  $Z$ , with coefficient vector  $\Gamma$ . The wage equation was estimated using U.S. Census Bureau CPS March survey data for years 1977-2001. We used the fitted values of this equation to impute wages to all workers in the data set. Using estimated wages and hours of individual workers, hours and imputed compensation are computed by industry and by four types of workers. The four worker types that we use are technology workers and three other worker types based on education attained (high school dropout, high school graduate, and

college plus).<sup>12</sup>

With these estimates from the CPS, we disaggregated the GPO employee hours and compensation paid, to obtain these variables by worker type consistent with the aggregates we observe in our augmented GPO dataset. We then aggregated hours of the four worker types by industry using Tornqvist compensation weights to obtain labor services. The labor quality index is defined as labor services divided by hours, and so, labor services are defined as labor quality times hours.

## **IV. Applications**

### *IV.A. Productivity Growth of Nonfarm Business*

As an initial exercise, we estimated total factor productivity (TFP) by industry and by legal form of organization, aggregated to private nonfarm business. At the individual industry level, we model the growth rate of TFP as the growth rate of real net output less the share-weighted growth rates of real intermediate inputs, labor input, and capital services. Net output is defined as gross output less the consumption of own-industry output as intermediate inputs; likewise, intermediate inputs are stripped of own-industry products. We use data from the input-output tables on the ratio of net output to gross output to estimate own-industry inputs. The data on real gross output, intermediate inputs, and cost-weighted expenditure shares come from our modified GPO dataset.

---

<sup>12</sup>For the years 1977-1982 ICT workers are defined as computer programmers, computer systems analysts, computer specialists, n.e.c., electrical and electronic engineers, and computer and peripheral equipment operators. For the years 1983-2000 ICT workers are defined as electrical and electronic engineers; computer systems analysts and scientists; operations and systems researchers and analysts; supervisors, computer equipment operators; chief communications operators; computer operators; and peripheral equipment operators.

To calculate aggregate TFP growth we take a weighted sum of the individual components, where the weights are calculated as sketched in Domar (1961).<sup>13</sup> We estimate the ratio of net output to gross output in each industry times the ratio of net output to gross output of all private industries excluding farm and owner-occupied housing as measured in the 1982, 1987 and 1992 input-output tables. We interpolate these ratios between years and then multiply them by the ratio of gross output in our dataset for each industry to gross output of all private nonfarm industries to obtain annual Domar weights. The contribution of inputs (excluding materials) and TFP to nonfarm business net output growth equals the weighted sum of the contributions to growth of the inputs and TFP to individual industry net output growth, where the weights are the annual Domar weights. The contribution from materials is calculated as the growth rate of net output less the sum of the contributions from the other inputs and TFP. As noted by Domar, the weighted sums of TFP growth rates measures the increase in aggregate output holding the factors of production constant, which is the closest thing to the concept of technical progress that we have.

Table 2 reports the growth rate of aggregate net output for private nonfarm businesses over each of the time periods considered, as well as an estimate of the contributions to growth from the use of materials, capital, labor, and total factor productivity. As described in the table, net output grew on average 3.1 percent per year. Capital services contributed 1.3 percentage points per year, and labor hours added 2/3 percentage point. We estimate that increases in the quality of labor contributed a little over 1/2 percentage point to net output growth. The Domar weighted average across industries of labor quality contributed 0.17 percentage point, while the

---

<sup>13</sup> See also Gollop (1987) and Hulten (1978) for a more detailed discussion of the derivation and

Domar-weighted average of the contribution of hours less the simple sum of NFB hours times labor share, which we refer to as reallocation effects, added 0.37 percentage point. These estimates, including the reallocation effects, are a little higher than is implied by the results in Jorgenson, Ho and Stiroh (2002) and in Aaronson and Sullivan (2001). We estimate that TFP rose on average 0.45 percent per year. Over the 1996-2001 period, net output climbed 3.9 percent per year. TFP accelerated to 0.9 percent per year, and the average contribution of high-tech capital services increased to 1.2 percentage points.

Tables 3a-3c report the disaggregated estimates for the sixty industries that we analyze over the three selected time periods. As has been noted elsewhere, although technical progress has always been rapid in the machinery manufacturing (which includes computers) and electrical machinery manufacturing industries (which includes communication equipment and semiconductors) these industries contributed importantly to the acceleration in TFP. Technical progress also picked up in the trade industries, as did the growth rate of their stock of high-tech equipment. Some other industries, such as depository institutions and business services, also pushed up their rates of investment in high-tech equipment. But, total factor productivity growth increased only 0.3 percentage point in depository institutions and fell sharply in business services.

Table 4 reports total factor productivity growth split between corporate and non-corporate private businesses. At the aggregate level, the acceleration noted in table 2 in nonfarm business TFP is due to the sharp improvement among noncorporate business. Indeed, among all major components, TFP rose more rapidly among noncorporate businesses than corporations.

---

interpretation of the Domar weights.

This could be an artifact of mismeasured capital services, however. As shown in the bottom half of the table, the contribution to growth from capital services was more rapid among corporations than noncorporate businesses.

#### *IV.B. Y2K*

In the late 1990's, businesses spent a large amount of money working to fix potential Y2K bugs. Software that could not recognize that the year represented by the two-digit number "00" was one year larger than the year "99" had to be modified or replaced. Industry reports indicate that some firms saw buying whole new systems, including hardware, as preventive maintenance.

These stories suggest that the rate of depreciation and discards of computers and software was unusually high in advance of the century data change. The models that we employ to estimate capital stocks do not directly measure this rate. Unless augmented, these models assume that the rate is a function of the stock and age of equipment of each vintage. As a small experiment with our system, we adjust the stocks of computers and software assuming that some share of Y2K spending represented additional scrappage.

To parameterize the experiment, we used figures reported by the U.S. Department of Commerce (1999). That report cites a study from International Data Corporation that public and private spending from 1995 to 2001 to fix the Y2K problem was roughly \$114 billion. It also cites an OMB report that the federal government spent a little over \$8 billion and a Federal Reserve study that suggests that spending by state and local governments was roughly half of federal spending. The Commerce report also provides some figures developed by Edward Yardeni of the distribution of spending across industries. We used the aggregate estimates to calculate baseline spending on Y2K by the private sector over 1995-2001, and we used the



Yardeni estimates to split them across broad industry aggregates. We assume that Y2K spending across different types of computer equipment and software was the same as total spending, except that we adjusted upwards the fraction on software by 50 percent based on some IDC figures on the split on spending between hardware and software to redress the Y2K bug.

Two considerations suggest these figures are not precise. IDC indicates that a lower and upper range for spending was plus or minus 50 percent. In addition, all of this Y2K spending does not necessarily reflect additional spending *on investment*. Estimates from IDC indicate that only 27 percent of worldwide spending was on “hardware or software”, whereas the rest was on “internal or external” spending, which may not have been counted as investment. As a lower bound, we assume none of the "internal or external" spending was investment; as an upper bound, we assume all of it was. This leaves a wide range of investment of \$14 to \$152 billion, which we assume also represents the additional scrapping of older stocks of hardware and software.

Table 5 reports the change in estimates of TFP by broad aggregates when one assumes that the upper bound of Y2K spending (\$150 billion) went to replacing high-tech equipment and software that was scrapped and replaced.<sup>14</sup> The largest effect on any aggregate in any year is +0.52 percentage point in the communications industry in 1997. The extra scrapping reduces the growth rate of capital services. Because real output is not changed, the lower contribution from capital services means that TFP must have been higher, in this case by 0.52 percentage point. In a few industries, such as communications, depository and nondepository institutions, and

---

<sup>14</sup> For the exercise, the underlying database was first aggregated to the level of detail available for the Y2K spending. For each of these activity groups, TFP was computed using net output and net intermediate use concepts. For higher level aggregates, the TFP was aggregated using appropriate Domar weighting.

business and miscellaneous professional services, the effect of Y2K scrappage could be important. For the rest, the effect appears to have been small relative to the average year-to-year variation in TFP. In total, if capital services are adjusted along the lines suggested above, the rate of growth in TFP would be 16 basis points higher in the last half of the 1990s and 13 points lower in 2000 and 2001. Assuming a more moderate level of Y2K spending that represents replacement investment (\$50 billion) reduces the cumulative effect to one-quarter of the upper-bound effect.

## **V. Conclusion**

This paper explicates a general approach to the problem of building a consistent data set for the study of economic issues. Coding observations in a relational database allows us to easily manipulate economic data, while the meta data help us to preserve the numerous linear relations across variables. The tools that we have developed take advantage of the standardized data and meta data in order to build a consistent data set.

The system was originally conceived to aid in the study of productivity. To that end, we started with the BEA's GPO data. We concorded the GPO data before 1987, which are organized using the 1972 SIC, to the more recent data, which use the 1987 SIC to classify industries. We then supplemented the dataset by including estimates of employee and all persons hours from the NIPAs and the BLS, as well as estimating some missing pieces of data, such as gross output for some industries before 1987 and some price deflators. We also concorded the BEA's estimates of investment by industry and by type to the GPO data. To study productivity, we linked data from the input-output tables to calculate Domar weights; we

incorporated data from the Current Population Survey, March Supplement 1977-2001, to estimate labor services; and, we employed some specialized tools that we developed to estimate capital stocks, capital services, and TFP. Finally, we decomposed all of the data by legal form of organization, controlling the estimates to be consistent with industry totals and aggregate legal form totals in the NIPAs.

Our overall estimates of TFP growth by industry point to the same qualitative results seen elsewhere. TFP accelerated in the last half of the 1990s and was particularly high in most industries outside of the service sector. The contribution to output growth from increased investment in high-tech capital equipment also picked up. We also demonstrated how the system could be employed to reconsider assumptions made in the construction of data and counterfactual exercises. In this small experiment, we took estimates of the amount of spending to remedy the Y2K problem and assumed that some fraction of this estimate was not an increment to the capital stock but instead purely replaced an unusually high amount of capital that was scrapped because it was potentially infected with the Y2K bug. Except for a few industries, the effects on TFP were likely small unless one were to assume that the scrappage associated with the century date change was very large.

A few obvious extensions appear possible. Fully incorporating the input-output data, including making them fully consistent with the value added data in the GPO dataset, would open up several research avenues. Immediately, it would allow us to have a fully consistent application of Domar weighting. It would allow us to study various price-markup models and to perform various counterfactuals, such as the effects of different productivity growth rates among intermediate producers on prices and aggregate productivity. If at the same time, separate

estimates of input-output tables at the same level of aggregation controlled to the current expenditure-side estimates of GDP were available, we could study the statistical discrepancy. Extending the input-output tables further back and incorporating auxiliary information on prices will enable us to estimate industry price deflators before 1977.

In putting together our preliminary estimates of capital services, we simply used the BEA's estimates of investment by industry and by type that it employs to estimate capital consumption and wealth. However, these estimates are based on limited data of investment by industries outside of census years and are not based on any systematic information on investment by both industry and by type in any year (see for instance, Bonds and Aylor, 1998). Indeed, even though the BEA has made these data publicly available on their website, they consider them unpublished because they do not rise to their usual standards for statistical reliability. In the future, we plan to examine how sensitive the capital services estimates are to other plausible distributions of investment. Based in part on conversations with our colleagues, we suspect that the distribution of computer investment could matter importantly, but for other types of equipment, the effects may be small. At the same time, we plan to examine how important the depreciation estimates are for estimates of capital services.

Finally, the system has the tools necessary to construct a research database starting with the most micro-level datasets. Many of the problems of switching classifications and cross classification would be better approached by working with plant and firm-level data. For example, a better concordance between the SIC and NAICS could be developed by attaching SIC and NAICS codes to each firm or establishment in a particular year (based on the same logic used to apply the original activity code to a respondent in the survey) and then tabulating a

concordance for each relevant variable. Indeed, a joint Federal Reserve-Census project is currently under way to develop such a concordances for manufacturing using the Longitudinal Research Database (Klimek and Bayard, 2002). The same method could be used in making a firm-establishment cross classification by linking enterprise, firm, and establishment codes at the micro level, and then merging and aggregating different data sources to create a cross-classification table.

## References

- Aaronson, Daniel and Daniel Sullivan (2001), "Growth in Worker Quality", *Federal Reserve Bank of Chicago Economic Perspectives* 25(4) (Fourth Quarter) 53-74.
- Bacharach, Michael (1965), "Estimating Nonnegative Matrices from Marginal Data", *International Economic Review* 6(3), (September) 294-310.
- Bartelsman, Eric J. and J. Joseph Beaulieu (2004), "Techniques to Reconcile Data with Linear Constraints", Mimeo, Federal Reserve Board.
- Bartelsman, Eric J. and Wayne Gray (1996), "The NBER Manufacturing Productivity Database", NBER Technical Working Paper #T0205 (October).
- Bell, William R. and David W. Wilcox (1993), "The Effect of Sampling Error on the Time Series Behavior of Consumption", *Journal of Econometrics*, 55(1-2) (January-February), 235-265.
- Bonds, Belinda and Tim Aylor (1998), "Investments in New Structures and Equipment in 1997 by Using Industries", *Survey of Current Business* 78(12), (December), 26-51.
- Codd, E.F. (1990), *The Relational Model for Database Management*, Reading, MA: Addison-Wesley.
- Corrado, Carol and Lawrence Slifman (1999), "Decomposition of Productivity and Unit Costs", *American Economic Review: Papers and Proceedings*, 89(2) (May), 328-32.
- Domar, Evsey D. (1961), "On the Measurement of Technological Change", *Economic Journal*, 71(284), 709-729.
- Fraumeni, Barbara M. (1997), "The Measurement of Depreciation in the U.S. National Income and ProductAccounts", *Survey of Current Business* 77(7) (July), 7-23.
- Golan, A., G. Judge, and S. Robinson (1994), "Recovering Information from Incomplete or Partial Multisectoral Economic Data", *Review of Economics and Statistics* 76(3), 541-549.
- Gollop, Frank M. (1987), "Modeling Aggregate Productivity Growth: The Importance of Intersectoral Transfer Prices and International Trade", *Review of Income and Wealth* 33(2) (June), 211-227.
- Griliches, Zvi (1986), "Economic Data Issues" in *Handbook of Econometrics*, Vol. III, Eds. Zvi Griliches and Michael D. Intriligator, North Holland: Oxford, England.
- Günlük-Şenesen, G. and J.M. Bates (1988), "Some Experiments with Methods of Adjusting Unbalanced Data Matrices", *Journal of Royal Statistical Society, Series A* 151(3), 473-490.
- Hall, Robert E. and Dale W. Jorgenson (1967), "Tax Policy and Investment Behavior", *American Economic Review* 57(3) (June), 391-414.
- Hulten, Charles R. (1978), "Growth Accounting with Intermediate Inputs", *Review of Economic*

- Studies* 45(3) (October), 511-518.
- Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh (2002), "Information Technology, Education, and the Sources of Economic Growth across U.S. Industries", mimeo Harvard University (April).
- Jorgenson, Dale W. and Kevin J. Stiroh (2000), "Raising the Speed Limit: U.S. Economic Growth in the Information Age", *Brookings Papers on Economic Activity*, 0(1), 125-211.
- Klimek, Shawn D. and Kimberly N. Bayard (2002), "Classifying the Census of Manufactures from the Standard Industry Classification System, 1963 to 1992", mimeo, Center for Economic Studies, U.S. Census Bureau.
- Miron, Jeffrey A. and Stephen P. Zeldes (1989), "Production, Sales, and the Change in Inventories: An Identity That Doesn't Add Up", *Journal of Monetary Economics* 24 (1) (July), 31-51.
- Mohr, Michael F. and Charles E. Gilbert (1996), "Capital Stock Estimates for Manufacturing Industries: Methods and data", Board of Governors of the Federal Reserve System (March) ([www.federalreserve.gov/releases/G17/capital\\_stock\\_doc-latest.pdf](http://www.federalreserve.gov/releases/G17/capital_stock_doc-latest.pdf)).
- Moylan, Carol (2001), "Estimation of Software in the U.S. National Accounts: New Developments", OECD, Statistics Directorate Working Paper STD/NA(2001) 25 (September, 24).
- Parikh, Ashok (1979), "Forecasts of Input-Output Matrices Using the R.A.S. Method", *Review of Economics and Statistics* 61(3) (August), 477-481.
- Parker, Robert and Bruce Grimm (2000), "Recognition of Business and Government Expenditures for Software as Investment: Methodology and Quantitative Impacts, 1959-98", the Bureau of Economic Analysis (<http://www.bea.gov/bea/papers/software.pdf>).
- Parker, Robert P. and Eugene P. Seskin (1997), "The Statistical Discrepancy", *Survey of Current Business* 77(8) (August), 19.
- Postner, Harry H. (1984), "New Developments Towards Resolving the Company-Establishment Problem" 30 (4) (December), 429-459.
- Schneider, M.H. and S.A. Zenios (1990), "A Comparative Study of Algorithms for Matrix Balancing", *Operations Research* 38, 439-455.
- Seskin, Eugene P. and Robert P. Parker (1998), "A Guide to the NIPA's", *Survey of Current Business* 78(3) (March), 26-68.
- Sinkhorn, R. (1964), "A Relationship between Arbitrary Positive Matrices and Doubly Stochastic Matrices", *Annals of Mathematical Statistics* 35, 876-879.
- Stone, R. and J.A.C. Brown (1962). *A Computable Model of Economic Growth*. London: Chapman and Hall.
- Stone, Richard, D.G. Champernowne, and J.E. Meade (1942), "The Precision of National Income Estimates", *The Review of Economic Studies* 9(2) (Summer), 111-125.

- U.S. Department of Commerce, Economics and Statistics Administration (1999), “The Economics of Y2K and the Impact on the United States”, November 17.
- Weale, Martin (1985), “Testing Linear Hypothesis on National Account Data”, *Review of Economics and Statistics*, 67(4) (November) 685-689.
- Wilcox, David W. (1992), “The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work”, *American Economic Review* 82(4) (September), 922-941.
- Yuskavage, Robert E. (2002), “Gross Domestic Product by Industry: A progress report on accelerated estimates”, *Survey of Current Business* 82(6) (June), 19-27



**Table 1**  
**Comparison of 2001 Compensation and Profits Data in the GPO**  
**and the NIPA Datasets\***

	<b>Compensation</b>		<b>Profits w/ IVA</b>	
	<b>GPO</b>	<b>NIPA</b>	<b>GPO</b>	<b>NIPA</b>
Manufacturing	939.2	939.2	52.1	83.4
Transportation and Utilities	382.1	382.1	23.8	27.7
Wholesale Trade	379.8	379.8	48.5	44.8
Retail Trade	531.1	531.1	79.3	79.1
Remaining domestic private industries	2586.9	2586.9	320.8	524.4

\*GPO and NIPA compensation data are collected on an establishment basis. NIPA profits data are collected by firms; the GPO converts these data to an establishment basis.

**Table 2**  
**Growth Accounting, Nonfarm Business**

	<b>Capital Services</b>				<b>Labor Services</b>			<b>Net</b>	
	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Reall.</b>	<b>Qual.</b>	<b>TFP</b>	<b>Output</b>
1978-2001	0.16	0.70	0.24	0.35	0.66	0.37	0.17	0.45	3.10
1978-1989	-0.08	0.52	0.22	0.46	0.81	0.54	0.20	0.20	2.88
1990-1995	0.71	0.59	0.15	0.24	0.37	-0.05	0.17	0.54	2.72
1996-2001	0.09	1.17	0.36	0.24	0.63	0.46	0.10	0.88	3.93

**Table 3a**  
**Growth Accounting**  
**(1977-1989)**

	<b>Domar</b>	<b>Capital Services</b>				<b>Labor Serv.</b>		<b>TFP</b>	<b>Output</b>
	<b>weight</b>	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Qual.</b>		
Agricultural services	0.7	-0.28	0.12	0.31	0.19	3.02	0.06	2.09	5.52
Metal mining	0.2	-0.85	0.02	-0.31	0.20	-0.69	0.09	2.33	0.79
Coal mining	0.8	1.14	0.01	-0.24	0.39	-1.07	0.13	2.54	2.90
Oil and gas extraction	4.4	0.27	0.14	0.13	1.22	0.25	0.04	-2.97	-0.92
Other mineral mining	0.4	-0.37	0.03	0.04	0.26	0.05	0.11	0.45	0.57
Construction	10.7	0.06	0.01	-0.09	0.02	1.17	0.11	-0.47	0.81
Lumber and wood	1.3	0.27	0.03	-0.16	0.04	0.09	0.21	1.08	1.56
Furniture and fixtures	0.9	1.57	0.04	0.03	0.06	0.36	0.13	0.28	2.48
Stone, clay, and glass	1.5	0.30	0.13	-0.13	-0.01	-0.16	0.08	0.30	0.50
Primary metalss	3.3	-0.24	0.02	-0.08	-0.04	-0.91	0.10	-0.11	-1.26
Fabricated metals	3.9	-0.16	0.07	0.11	0.04	-0.20	0.10	0.54	0.50
Machinery	5.6	1.46	0.29	0.14	0.07	0.05	0.19	2.91	5.11
Electrical machinery	3.9	1.34	0.43	0.23	0.12	0.35	0.31	1.93	4.70
Motor vehicles	4.3	1.53	0.07	0.00	0.04	-0.31	0.14	-0.56	0.91
Other transportation equip.	2.7	2.01	0.33	0.07	0.09	1.02	0.28	-0.11	3.68
Instruments	2.6	2.99	0.28	0.04	0.12	0.50	0.26	1.19	5.37
Miscellaneous manufacturing	0.9	-0.70	0.06	-0.01	0.03	0.05	0.21	0.97	0.60
Food	7.7	1.30	0.06	0.07	0.02	0.02	0.05	0.72	2.24
Tobacco	0.5	5.50	0.10	0.25	0.18	-0.41	0.05	-6.35	-0.68
Textiles	1.3	-0.59	0.06	-0.10	-0.04	-0.58	0.12	1.36	0.23
Apparel	1.6	-0.12	0.04	-0.03	0.03	-0.50	0.13	1.24	0.78
Paper	2.2	1.74	0.12	0.31	0.05	0.03	0.15	0.11	2.51
Printing	2.9	2.35	0.30	0.08	0.09	1.20	0.14	-0.83	3.34
Chemicals	4.9	1.06	0.13	0.11	0.09	0.00	0.08	0.68	2.14
Petroleum	5.0	-1.06	0.03	0.04	0.01	-0.09	0.02	0.90	-0.16
Rubber and plastics	2.1	1.50	0.06	0.06	0.07	0.38	0.13	1.35	3.54
Leather	0.3	-3.47	0.01	-0.08	-0.02	-1.97	0.20	1.18	-4.15
Railroad transportation	0.9	0.33	0.09	-0.23	-0.51	-1.08	0.09	3.54	2.23
Local & interurban transit	0.4	-0.45	0.14	0.02	0.17	0.91	0.15	-1.77	-0.84
Trucking and warehousing	2.4	2.74	0.04	0.45	0.03	0.49	0.15	0.40	4.31
Water transportation	0.7	-1.37	0.03	-0.50	0.09	0.14	0.11	-0.10	-1.59

**Table 3a**  
**Growth Accounting**  
**(1977-1989)**

	<b>Domar</b>	<b>Capital Services</b>				<b>Labor Serv.</b>		<b>TFP</b>	<b>Output</b>
	<b>weight</b>	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Qual.</b>		
Transportation by air	1.6	3.09	0.13	0.44	0.17	2.21	-0.01	1.16	7.20
Pipelines, ex. natural gas	0.3	1.03	0.03	-0.03	-0.52	0.07	0.05	-1.09	-0.47
Transportation services	0.5	-0.36	0.34	0.03	0.07	2.50	0.20	0.64	3.42
Telephone and telegraph	3.3	0.34	1.44	0.19	0.86	0.05	0.11	1.56	4.56
Radio and television	0.9	2.53	1.01	0.13	0.88	0.16	-0.07	-1.44	3.20
Utilities	5.8	0.75	0.44	0.71	0.65	0.29	0.08	-0.89	2.04
Wholesale trade	11.8	0.21	0.30	0.16	0.29	1.04	0.13	1.50	3.63
Retail trade	16.2	0.89	0.51	0.16	0.19	1.80	0.12	-0.55	3.14
Depository institutions	4.5	0.98	0.97	0.50	0.51	0.58	0.13	-0.76	2.91
Nondepository institutions	0.6	0.30	1.76	1.53	0.08	15.94	0.20	-14.53	5.29
Security/commodity brokers	1.0	3.34	0.87	0.29	0.57	4.86	0.36	0.50	10.79
Insurance carriers	3.0	3.46	0.60	0.28	0.34	1.05	0.10	-3.45	2.37
Insurance agents	1.1	0.33	0.13	-0.05	0.02	2.12	0.15	-0.88	1.84
Nonfarm housing services	2.7	-0.03	0.00	0.00	1.62	-0.05	0.02	1.40	2.97
Other real estate	4.7	3.58	0.13	0.27	1.80	0.32	0.06	0.52	6.68
Investment offices	0.4	5.67	1.45	0.54	1.07	2.70	0.19	-8.14	3.48
Hotels and other lodging	1.4	1.70	0.07	0.16	0.56	1.00	0.15	-1.09	2.55
Personal services	1.1	1.01	-0.03	-0.04	0.05	1.24	0.16	-0.43	1.97
Business services	4.0	2.88	1.90	-0.24	0.14	5.01	0.32	-1.33	8.69
Automobile services	1.6	1.34	0.12	0.50	0.08	1.50	0.09	-0.12	3.52
Misc. repair services	0.7	1.43	0.08	-0.00	0.05	0.92	0.12	0.14	2.73
Motion pictures	0.4	0.81	0.16	0.35	0.25	2.37	-0.04	1.64	5.55
Amusement and recreation	0.9	0.76	-0.06	-0.15	0.21	0.72	-0.03	1.69	3.14
Health services	5.5	3.30	0.21	0.07	0.22	1.96	0.19	-2.08	3.86
Legal services	1.6	1.77	0.35	0.17	0.10	3.59	0.09	-2.18	3.88
Educational services	0.7	3.40	0.05	-0.00	0.12	0.20	0.02	0.04	3.83
Social services	0.5	7.63	0.17	0.04	0.05	2.88	0.00	-1.07	9.70
Membership organizations	0.8	0.34	0.06	0.01	0.02	4.65	0.12	-3.87	1.33
Other services	3.5	2.22	0.46	0.09	0.13	-2.40	0.15	5.51	6.14

**Table 3b**  
**Growth Accounting**  
**(1990-1995)**

	<b>Domar</b>	<b>Capital Services</b>				<b>Labor Serv.</b>			<b>Output</b>
	<b>weight</b>	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Qual.</b>	<b>TFP</b>	
Agricultural services	0.8	0.91	0.22	0.99	0.12	1.52	0.01	-0.80	2.97
Metal mining	0.1	-3.31	0.58	0.12	-0.27	-0.30	0.06	1.91	-1.21
Coal mining	0.5	-0.97	0.15	-0.07	0.14	-1.08	0.07	3.16	1.39
Oil and gas extraction	1.8	-0.94	0.09	-0.56	-0.27	-0.25	0.03	0.85	-1.03
Other mineral mining	0.3	-0.03	0.19	0.07	-0.07	-0.02	0.07	0.66	0.88
Construction	8.5	-0.11	0.11	0.00	0.01	0.06	0.13	-0.23	-0.02
Lumber and wood	1.2	1.86	0.10	-0.11	0.01	0.21	0.05	-1.73	0.40
Furniture and fixtures	0.8	1.31	0.10	0.02	0.02	-0.14	0.11	0.38	1.79
Stone, clay, and glass	1.1	-0.13	0.02	-0.10	-0.04	-0.19	0.15	1.02	0.74
Primary metalss	2.2	0.36	0.03	-0.09	-0.07	-0.33	0.06	1.25	1.20
Fabricated metals	3.0	0.74	0.15	0.02	0.01	0.05	0.06	0.80	1.84
Machinery	4.6	4.40	0.32	0.02	0.03	-0.09	0.05	2.02	6.74
Electrical machinery	3.7	4.19	0.43	0.24	0.09	-0.49	0.12	5.64	10.22
Motor vehicles	4.0	2.02	0.10	0.09	0.03	0.74	0.06	0.70	3.74
Other transportation equip.	2.3	-1.49	-0.03	-0.07	0.05	-2.38	0.05	-0.94	-4.80
Instruments	2.4	2.58	0.45	-0.03	0.06	-1.00	0.27	-1.07	1.25
Miscellaneous manufacturing	0.7	1.26	0.13	0.01	0.02	0.00	0.41	-0.47	1.35
Food	6.4	0.67	0.08	0.09	0.03	0.17	0.05	0.63	1.72
Tobacco	0.6	2.09	0.05	-0.08	-0.00	-0.38	0.03	-0.67	1.04
Textiles	1.0	0.58	0.20	-0.06	-0.03	-0.51	0.18	1.49	1.85
Apparel	1.2	1.48	0.11	-0.01	0.02	-0.70	0.19	0.33	1.41
Paper	2.1	1.59	0.16	0.21	0.04	0.02	0.12	-0.50	1.64
Printing	2.8	1.06	0.40	-0.03	0.05	0.21	0.13	-2.37	-0.55
Chemicals	4.6	-0.40	0.47	0.15	0.11	-0.07	0.11	0.79	1.17
Petroleum	2.6	0.40	0.20	0.17	-0.05	-0.14	0.03	-0.40	0.22
Rubber and plastics	2.1	2.56	0.14	0.24	0.06	0.47	0.11	0.87	4.45
Leather	0.1	-3.53	0.08	-0.13	-0.05	-2.20	0.20	2.52	-3.11
Railroad transportation	0.6	0.21	0.07	-0.21	-0.64	-0.32	0.06	3.99	3.15
Local & interurban transit	0.4	0.96	-0.01	0.15	0.24	1.03	0.12	-1.46	1.02
Trucking and warehousing	2.6	2.41	0.20	0.16	0.06	0.40	0.09	1.36	4.67
Water transportation	0.5	0.74	0.02	-0.65	0.07	0.38	0.04	1.60	2.20

**Table 3b  
Growth Accounting  
(1990-1995)**

	<b>Domar</b>	<b>Capital Services</b>				<b>Labor Serv.</b>		<b>TFP</b>	<b>Output</b>
	<b>weight</b>	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Qual.</b>		
Transportation by air	1.7	-0.93	0.31	0.17	0.07	3.16	0.18	0.10	3.07
Pipelines, ex. natural gas	0.1	1.31	0.69	-0.11	-0.38	-0.22	0.01	-3.02	-1.71
Transportation services	0.5	3.09	1.26	-0.18	0.12	1.74	0.07	-0.59	5.50
Telephone and telegraph	3.3	2.76	1.04	0.40	0.34	0.21	0.17	1.85	6.76
Radio and television	0.9	-4.50	1.69	0.22	0.86	0.24	-0.12	2.50	0.89
Utilities	5.1	0.25	0.28	0.34	0.42	-0.02	0.01	0.52	1.80
Wholesale trade	11.5	2.11	0.23	0.20	0.33	0.18	0.02	0.98	4.04
Retail trade	16.2	1.50	0.73	0.01	0.10	1.18	0.11	-0.92	2.71
Depository institutions	4.9	1.11	1.07	-0.34	0.34	-0.69	0.06	-0.16	1.39
Nondepository institutions	1.0	4.34	3.60	1.23	0.06	-2.73	0.04	2.59	9.13
Security/commodity brokers	1.5	3.86	0.10	0.46	0.30	2.54	-0.09	3.61	10.77
Insurance carriers	3.9	-0.62	0.81	0.21	0.27	-0.52	0.11	-0.11	0.15
Insurance agents	1.2	1.11	0.17	0.13	0.02	0.56	0.19	-3.52	-1.35
Nonfarm housing services	3.0	-0.49	0.00	0.00	0.57	0.23	0.01	0.68	1.00
Other real estate	6.8	2.49	0.55	0.01	0.80	-0.11	0.01	-0.14	3.61
Investment offices	0.4	1.09	0.01	-0.11	0.90	-0.05	-0.05	-0.24	1.56
Hotels and other lodging	1.6	0.93	0.10	0.16	0.31	-0.65	-0.01	1.29	2.13
Personal services	1.2	1.59	0.23	-0.03	0.05	0.50	0.23	-1.01	1.55
Business services	6.2	3.21	0.44	0.24	0.05	2.31	0.27	0.31	6.84
Automobile services	1.8	1.57	0.09	1.56	0.05	0.33	0.04	-1.20	2.43
Misc. repair services	0.7	3.88	0.26	-0.06	0.03	0.17	0.05	-1.72	2.61
Motion pictures	0.6	4.93	0.98	0.49	0.25	2.06	-0.02	-3.44	5.25
Amusement and recreation	1.2	4.61	0.05	0.20	0.23	0.21	-0.01	1.85	7.15
Health services	7.4	2.78	0.28	0.03	0.12	0.71	0.23	-1.54	2.61
Legal services	2.1	-0.05	0.18	-0.07	0.02	-0.21	0.43	-0.75	-0.44
Educational services	0.8	2.70	0.07	-0.05	0.07	-0.03	0.03	0.42	3.22
Social services	0.9	2.64	0.16	0.01	0.03	0.73	0.09	0.71	4.38
Membership organizations	0.8	2.56	0.08	-0.01	0.02	-18.85	0.18	19.84	3.81
Other services	4.2	0.83	0.23	-0.01	0.06	0.94	0.39	-0.24	2.18

**Table 3c**  
**Growth Accounting**  
**(1996-2001)**

	<b>Domar</b>	<b>Capital Services</b>				<b>Labor Serv.</b>		<b>TFP</b>	<b>Output</b>
	<b>weight</b>	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Qual.</b>		
Agricultural services	0.8	0.99	0.27	1.45	0.06	1.73	0.05	-0.10	4.45
Metal mining	0.1	-5.07	0.06	-0.19	-0.69	-1.58	-0.05	6.69	-0.83
Coal mining	0.3	-0.98	0.16	0.53	-0.02	-0.85	-0.04	2.48	1.29
Oil and gas extraction	1.6	2.23	0.26	0.01	0.02	0.01	-0.01	-1.61	0.90
Other mineral mining	0.2	-1.26	0.37	1.14	0.00	0.02	-0.06	1.67	1.88
Construction	8.5	0.79	0.14	0.29	0.01	1.93	-0.06	-0.27	2.84
Lumber and wood	1.1	1.32	0.12	0.11	-0.00	-0.26	0.13	-0.82	0.59
Furniture and fixtures	0.8	1.06	0.14	0.09	0.01	0.03	0.00	0.15	1.48
Stone, clay, and glass	1.1	0.95	0.27	0.53	0.04	0.16	0.03	-0.82	1.15
Primary metals	1.8	-1.60	0.08	-0.03	-0.05	-0.46	0.18	0.44	-1.45
Fabricated metals	2.8	1.40	0.20	0.07	0.01	-0.05	0.08	-0.27	1.44
Machinery	4.6	1.49	0.61	0.13	0.03	-0.47	0.07	3.67	5.52
Electrical machinery	4.0	3.67	0.62	0.31	0.18	-0.12	0.24	6.42	11.33
Motor vehicles	4.1	1.45	0.10	0.15	0.01	-0.17	0.06	0.02	1.62
Other transportation equip.	1.8	1.99	0.29	-0.00	0.01	-0.13	0.01	1.64	3.79
Instruments	2.0	3.25	0.63	0.04	0.05	-0.06	0.05	-1.27	2.68
Miscellaneous manufacturing	0.7	-0.13	0.16	0.07	-0.01	-0.50	0.06	2.00	1.65
Food	5.3	2.36	0.11	0.15	0.02	-0.04	-0.02	-1.53	1.07
Tobacco	0.6	5.78	0.05	-0.11	0.02	-0.25	-0.01	-9.16	-3.67
Textiles	0.7	-1.01	0.13	-0.04	-0.05	-1.87	0.10	0.01	-2.73
Apparel	0.9	-0.51	0.09	0.01	-0.00	-2.67	0.10	1.09	-1.89
Paper	1.7	-0.74	0.13	0.08	0.01	-0.43	0.03	-0.15	-1.06
Printing	2.5	0.42	0.68	0.05	0.01	-0.62	0.02	-1.00	-0.43
Chemicals	4.1	0.66	0.38	0.16	0.09	-0.14	0.09	-0.10	1.14
Petroleum	2.2	0.92	0.02	-0.11	-0.05	-0.14	0.01	0.20	0.85
Rubber and plastics	2.0	1.47	0.19	0.35	0.05	-0.18	0.10	0.62	2.61
Leather	0.1	1.92	0.21	0.01	-0.06	-3.91	0.16	-1.93	-3.60
Railroad transportation	0.5	-0.19	0.14	0.02	-0.47	-0.54	0.03	1.26	0.25
Local & interurban transit	0.4	-0.57	0.13	0.85	0.17	1.42	0.04	0.11	2.15
Trucking and warehousing	2.7	1.66	0.15	0.31	0.12	0.86	0.00	-0.22	2.88
Water transportation	0.5	1.28	0.07	-0.24	0.04	0.27	0.03	1.01	2.47

**Table 3c**  
**Growth Accounting**  
**(1996-2001)**

	<b>Domar</b>	<b>Capital Services</b>				<b>Labor Serv.</b>			<b>Output</b>
	<b>weight</b>	<b>Mat.</b>	<b>ICT</b>	<b>Eqp.</b>	<b>Str.</b>	<b>Hours</b>	<b>Qual.</b>	<b>TFP</b>	
Transportation by air	1.7	0.89	0.84	0.99	0.07	0.32	0.05	0.13	3.29
Pipelines, ex. natural gas	0.1	-1.81	0.77	0.18	-0.35	-0.16	0.01	0.74	-0.61
Transportation services	0.6	1.69	2.48	0.50	0.07	0.68	0.05	-0.64	4.84
Telephone and telegraph	4.0	5.93	2.56	0.51	0.63	0.93	0.08	1.78	12.40
Radio and television	1.1	4.95	3.44	0.73	0.64	1.36	-0.05	-6.92	4.14
Utilities	4.3	1.43	0.25	0.16	0.44	-0.19	0.02	-1.04	1.06
Wholesale trade	11.1	-0.55	0.53	0.22	0.22	0.21	0.08	4.00	4.73
Retail trade	15.8	0.56	1.36	0.23	0.08	0.74	0.03	1.82	4.82
Depository institutions	5.2	0.40	1.84	-0.15	0.14	-0.01	0.13	0.28	2.62
Nondepository institutions	1.7	0.89	5.51	1.84	0.04	1.59	0.07	-1.39	8.55
Security/commodity brokers	2.7	1.49	0.62	0.47	0.18	3.41	0.26	7.66	14.08
Insurance carriers	3.8	-1.38	1.15	0.18	0.18	0.60	0.04	-1.28	-0.51
Insurance agents	1.2	1.76	0.43	0.34	0.02	0.74	0.07	-1.14	2.23
Nonfarm housing services	2.6	0.34	0.00	0.00	1.12	0.12	0.00	-0.80	0.77
Other real estate	7.2	1.98	0.77	0.22	0.55	0.17	0.00	1.38	5.07
Investment offices	0.5	1.32	1.47	0.72	1.04	0.33	0.15	6.51	11.54
Hotels and other lodging	1.6	0.73	0.19	0.21	0.54	1.84	0.14	-2.14	1.52
Personal services	1.2	1.23	0.16	0.05	0.06	0.63	0.11	-0.11	2.13
Business services	8.6	5.06	1.70	0.40	0.07	3.19	0.25	-1.57	9.10
Automobile services	1.8	1.00	0.17	1.05	0.02	0.81	0.07	0.62	3.73
Misc. repair services	0.7	2.25	0.41	0.18	0.02	-0.19	0.07	-2.49	0.25
Motion pictures	0.6	-0.02	0.71	0.22	0.27	1.35	0.12	-0.31	2.34
Amusement and recreation	1.4	2.29	0.18	0.37	0.37	1.73	0.09	-2.25	2.78
Health services	7.3	2.39	0.38	0.07	0.11	1.13	0.03	-1.44	2.66
Legal services	2.0	1.37	0.53	0.02	0.02	1.27	0.12	-0.48	2.85
Educational services	0.8	2.65	0.19	-0.01	0.10	0.27	0.00	-0.48	2.71
Social services	1.0	5.27	0.29	0.02	0.03	0.64	0.02	-0.32	5.95
Membership organizations	0.8	2.41	0.25	0.01	0.03	0.99	0.07	-1.13	2.63
Other services	5.0	2.61	0.54	0.03	0.05	3.17	-0.03	0.51	6.88

**Table 4**  
**Output Contribution from**  
**Capital Services and TFP by Legal Form of Organization**

	1977-1989		1990-1995		1996-2001	
	Corp.	NCrp.	Corp.	NCrp.	Corp.	NCrp.
<b>TFP</b>						
Nonfarm private business	0.30	-0.06	0.78	-0.11	0.64	1.77
Agricultural services	3.11	1.35	-0.92	-0.44	-0.52	0.95
Mining	-1.43	-2.64	0.87	4.14	-0.42	0.73
Manufacturing	0.89	2.24	0.93	-0.41	0.72	2.71
Transportation & utilities	0.18	1.12	0.88	3.52	-0.41	0.83
Wholesale trade	1.52	-0.07	1.19	-1.07	3.97	3.29
Retail trade	-0.91	-0.32	-0.87	-1.60	1.76	2.15
Finance, insurance & real estate	-2.21	1.65	0.25	0.50	1.06	2.28
Services	0.76	-1.33	0.64	-0.72	-1.99	-0.14
	1977-1989		1990-1995		1996-2001	
	Corp.	NCrp.	Corp.	NCrp.	Corp.	NCrp.
<b>Capital Services</b>						
Nonfarm private business	1.18	1.27	0.99	0.43	1.83	0.85
Agricultural services	-4.15	-0.89	2.02	0.26	2.58	0.36
Mining	1.13	2.00	-0.29	-1.06	0.67	-0.97
Manufacturing	0.38	0.06	0.40	-0.11	0.59	0.12
Transportation & utilities	1.39	1.31	1.07	0.52	2.06	1.23
Wholesale trade	0.79	-2.16	0.68	0.81	0.93	0.98
Retail trade	0.95	-0.29	0.86	0.44	1.72	0.96
Finance, insurance & real estate	2.24	1.63	1.61	0.53	2.59	0.75
Services	1.01	0.44	0.71	0.16	1.39	0.41



**Table 5**  
**Effect of Y2K Spending on TFP Growth**

	\$150 billion							\$50 b.	
	1995	1996	1997	1998	1999	2000	2001	Cum.	Cum.
Nonfarm private business	0.05	0.17	0.26	0.22	0.10	-0.12	-0.13	0.56	0.14
Forestry, fishing, agrc. services	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	0.01	0.00
Mining and construction	0.02	0.05	0.09	0.08	0.04	-0.03	-0.04	0.21	0.05
Manufacturing	0.02	0.06	0.09	0.08	0.03	-0.04	-0.04	0.21	0.05
Durable goods	0.02	0.08	0.11	0.10	0.04	-0.05	-0.05	0.25	0.06
Electronic eq. and instr.	0.04	0.14	0.17	0.13	0.05	-0.07	-0.08	0.39	0.10
Motor vehicles & equipment	0.01	0.02	0.02	0.02	0.01	-0.01	-0.01	0.06	0.01
Other mfg. durables	0.02	0.06	0.10	0.09	0.04	-0.04	-0.04	0.21	0.05
Nondurable goods	0.01	0.04	0.06	0.05	0.02	-0.02	-0.03	0.13	0.03
Chemical, petroleum, coal	0.01	0.03	0.04	0.04	0.02	-0.02	-0.02	0.10	0.02
Excluding petrochemicals	0.01	0.04	0.06	0.05	0.02	-0.02	-0.03	0.13	0.03
Transportation and utilities	0.05	0.17	0.25	0.21	0.11	-0.11	-0.11	0.57	0.14
Communications	0.11	0.36	0.52	0.41	0.21	-0.23	-0.21	1.16	0.29
Excluding communications	0.03	0.08	0.13	0.12	0.06	-0.05	-0.06	0.31	0.08
Wholesale and retail trade	0.04	0.12	0.18	0.16	0.08	-0.08	-0.09	0.41	0.10
Finance, insurance & real estate	0.05	0.14	0.21	0.17	0.07	-0.09	-0.10	0.45	0.11
Depository & nondepository	0.09	0.32	0.44	0.35	0.14	-0.19	-0.21	0.93	0.23
Other finance & insurance	0.05	0.09	0.14	0.12	0.05	-0.06	-0.07	0.31	0.08
Real estate	0.01	0.04	0.07	0.06	0.03	-0.02	-0.03	0.17	0.04
Services	0.06	0.20	0.29	0.24	0.09	-0.15	-0.14	0.59	0.15
Business & other services	0.08	0.25	0.36	0.29	0.09	-0.20	-0.18	0.69	0.17
Recreation & motion pictures	0.03	0.09	0.13	0.10	0.00	-0.06	-0.06	0.24	0.06
Other services	0.05	0.16	0.24	0.20	0.09	-0.10	-0.11	0.52	0.13

## Appendix

### Concording the Input-Output Tables to the GPO data

A handful of input-output industries had to be split among two or more GPO industries. The tables below describe how the weights for the concordance were calculated in order to allocate the outputs and inputs of these IO commodities and industries among the GPO industries. The 1982 table was mapped to 1972 GPO industries and then concorded to 1987 industries using the same concordance that was used for gross output in the GPO. In calculating price deflators, the reverse was done, and the 1987 table was concorded to the 1972 SIC.

After the concordance, the I-O tables were adjusted to account for the new treatment of software in the NIPAs. All three tables (1982, 1987, 1992) treat pre-packaged and custom software as an intermediate input and do not count own-account software as an output. As of the 2000 revision, the BEA began to count software as investment (Parker and Grimm, 2000). To adjust the I-O tables, we reduced the amount of the use of the commodity “computer and data processing services” by the amount of investment in pre-packaged and custom software, and we raised the make of the same commodity by the amount of own-account software investment.<sup>15</sup>

The first columns of tables A1-A3 report the IO code, and the second columns indicate to which GPO industries these IO codes map. The next three columns show how, in one of two ways, the weights were calculated. Either the weight was written down directly, or it was set as some fraction of a particular indicator. If the weights were entered directly, the column **Indc.** equals “Dir”; the column **Nmb.** reports the value of the weight in billions of dollars; and the last column reports the source for the weight. Otherwise, the weight equals the value in **Nmb.** times the indicator noted in the columns **Indc.** and **GPO indc.** The values in the **Indc.** column can equal GO (gross output), GP (gross product), or Sh (manufacturing shipments). The column **GPO indc.** reports the particular industry that is used as an indicator. If **Nmb.** does not equal one, the **Comment** column describes how the fraction was calculated.

For instance, the 1982 IO industry 11.0101 had to be split in two. The weight used to

---

<sup>15</sup> We did not adjust manufacturing in 1992 for custom software because Moylan (2001) indicates that the 1992 and 1997 censuses did not collect information on purchases of services by manufacturers, which we take to mean what is now known as custom software investment.

calculate the fraction that is part of GPO industry 15-17 was set to 0.796 times the gross product of GPO industry 15-17; the weight used to allocate the rest of 11.0101 to GPO industry 65re was set equal to 0.122 times the gross product of industry 65re.

**Table A1**  
**Splitting IO industries to different GPO industries, 1982**

<b>IO code</b>	<b>GPO Indus.</b>	<b>Indc.</b>	<b>GPO indc.</b>	<b>Nmb.</b>	<b>SIC</b>	<b>Comment</b>
04.0001	01-2	Dir		.5	0254, 0279pt	No information so split 04.0001 evenly between 01-2 and 07-09.
	07-9	Dir		.5	071-2,5-6, 085, 092	
11.0101	15-17	GP	15-17	.796	15, 16	Ratio of empl. of 15 & 17 to 15-17 in 1982. Ratio of empl. in 655 to 65 times ½ to split 655 b/n 6552 and 6551.
	65re	GP	65re	.122	6552	
11.0103	15-17	GO	15-17	1.0	15-17	Ratio of empl. in 655 to 65 in 1982 times ½ to split 655 b/n 6552 and 6551.
	65re	GO	65re	.122	6552	
11.0602	10	GO	1081	1.0	1081	
	11-12	GO	1112	1.0	1112	
	13	GO	138	1.0	138	
	14	GO	1481	1.0	1481	
11.0603	10	GO	1081	1.0	1081	
	11-12	GO	1112	1.0	1112	
	14	GO	1481	1.0	1481	
14.1801	20	Sh	2051	1.0	2051	Ratio of empl. in 5462 (bakeries) to empl. in all food stores excluding grocery stores in 1982.
	52-59	GO	542-9	.434	5462	
18.0400	23	Sh	231-8	1.0	231-8	Shipments of 39996 (Furs dressed and dyed) in 1982 Census.
	39	Dir		.1	3999pt	
38.0400	28	Sh	2819	1.0	2819	
	33		3334	1.0	3334	
65.0100	40	GP	40	1.0	40	Have to use value added because no gross output data are available for 47. Assume 4741, 4738, 4785, & 4789 are same size and 4789 split evenly between 65.0100 and 65.0300.
	47	GP	47	.05	474, 4789pt	
65.0300	42	GP	42	1.0	42	Have to use value added because no gross output data available for 47. Same as with 65.0100.
	47		47	.025	4789pt	
69.0200	52-59	Clc	Mixed		52-7,9 ex. 5462	Calculated as GP(52-9)*(1-GO(548)) / GO(52-9) Assumed to be small Revenue of 8042 from 1982 Census.
	73	GP	73	.01	7396	
	80	Dir		14.5	8042	

**Table A1, continued**  
**Splitting IO industries to different GPO industries, 1982**

<b>IO code</b>	<b>GPO indus.</b>	<b>Indc.</b>	<b>GPO indc.</b>	<b>Nmb.</b>	<b>SIC</b>	<b>Comment</b>
70.0200	61 67	GP GP	61 67	1.0 .888	61 67 ex. 6732	1-ratio of empl. of 673 in 2000 (from Occupation by industry data) to empl. in 67 times ½ to split b/n 6732 and 6733.
77.0302	07-09 80	GO Dir	07	.140	074 8049, 807-9	Ratio of empl. in 074 to 07 in 1982. Revenue of 8049 and 807-9 from 1982 Census.
77.0504	67  84, 89  86	GP  GP	67  84, 89  86	.1125  .073  .083	6732  84, 8922  865, 9	Ratio of empl. of 673 in 2000 (from Occupation by industry data) to empl. in 67 times ½ to split b/n 6732 and 6733.  ¼ of empl. in 873 + empl. in 84 in 1999 (from Occupation data) divided by empl. in 84, 87, and 89.  Ratio of empl. in political organizations and membership organizations, n.e.c. to all empl. in 86 (from Occupation data).

**Table A2**  
**Splitting IO industries to different GPO industries, 1987**

IO code	GPO Indus.	Indc.	GPO indc.	Nmb.	SIC	Comment
04.0001	01-2	Dir		.5	0254,	No information so split 04.0001 evenly between 01-2 and 07-09.
	07-09			.5	0279pt 071-2, 075-6, 085, 092	
11.0000	15-17	GO	15T7	1.0	15-17	Ratio of empl. in 655 to 653 in 1987 times ½ to split 655 b/n 6552 and 6551.
	65re	GO	653	.149	6552	
11.0602	10	GO	1081	1.0	1081	
	12	GO	1241	1.0	1241	
	13	GO	138	1.0	138	
	14	GO	1481	1.0	1481	
11.0603	10	GO	1081	1.0	1081	
	12	GO	1241	1.0	1241	
	14	GO	1481	1.0	1481	
14.1801	18	Sh	2051	1.0	2051	Ratio of empl. in 5461 (bakeries) to empl. in all food stores excl. grocery stores in 1987.
	52-59	GO	542-9	.485	5461	
38.0400	28	Sh	2819	1.0	2819	
	33		3334	1.0	3334	
65.0100	40	G0	40	1.0	40	Assume 4741, 4738, 4785, & 4789 are same size and 4789 split evenly between 65.0100 and 65.0300.
	47	G0	474-8	.375	4741, 4789pt	
65.0300	42	GO	42	1.0	42	Like 65.0100
	47		474-8	.125	4789pt	
69.0200	52-59	GO	527-9	1.0	52-7,9	Have to exclude 5462, as calculated above (14.1801) Revenue of 7396 from 1987 Census Revenue of 8042 from 1982 Census.
	52-59	GO	542-9	-.485	Ex. 5462	
	73	Dir		.3	7396	
	80	Dir		3.5	8042	
70.0200	61	GO	61	1.0	61	One minus ratio of empl. of 673 in 2000 (from Occupation by industry data) to empl. in 67 times ½ to split b/n 6732 and 6733.
	67	GO	67	.888	67 ex. 6732	
77.0302	07-09	GO	074	1.0	074	Revenue of 8043 and 8049 in 1987 Census
	80	Dir		3.6	8043, 8049	
	80	GO	807-9	1.0	807-9	
77.0504	67	GO	67	.112	6732	Ratio of empl. of 673 in 2000 (from Occupation by industry data) to empl. in 67 times ½ to split b/n 6732 and 6733.
	84	GO	84	1.0	84	
	86	GO	865	1.0	865	
	86	GO	869	1.0	869	

**Table A3**  
**Splitting IO industries to different GPO industries, 1992**

<b>IO code</b>	<b>GPO indus.</b>	<b>Indc.</b>	<b>GPO indc.</b>	<b>Nmb.</b>	<b>SIC</b>	<b>Comment</b>
04.0001	01-02	Dir		.5	0254, 0279pt	No information so split 04.0001 evenly between 01-2 and 07-09.
	07-09			.5	071-2, 075-6, 085, 092	
11.0101	15-17 63re	GO Dir	110101	1.0 4.6	15, 17 6552	Half of revenue of 6552 in 1992 Census.
11.0108	15-17 65re	GO Dir	110108	1.0 4.6	15, 17 6552	Half of revenue of 6552 in 1992 Census.
11.0602	10	GO	1081	1.0	1081	
	12	GO	1241	1.0	1241	
	13	GO	138	1.0	138	
	14	GO	1481	1.0	1481	
11.0603	10	GO	1081	1.0	1081	
	12	GO	1241	1.0	1241	
	14	GO	1481	1.0	1481	
65.0100	40	GO	40	1.0	40	
	47	Dir		1.9	474	Revenue of 474 in 1992 Census
70.0200	61	GO	61	1.0	61	
	67	GO	67	.888	67 ex. 6732	One minus ratio of empl. of 673 in 2000 (from Occupation by industry data) to empl. in 67 times ½ to split b/n 6732 and 6733.
77.0504	67	GO	67	.112	6732	Ratio of empl. of 673 in 2000 (from Occupation by industry data) to empl. in 67 times ½ to split b/n 6732 and 6733.
	84	GO	84	1.0	84	
	86	GO	865	1.0	865	
	86	GO	869	1.0	869	