PRPM 7|29|02
BARTELSMAN/BEAULIEU
9:00AM

PRELIMINARY
PLEASE DO NOT QUOTE
EMPIRICAL RESULTS

A Consistent Accounting of U.S. Productivity Growth*

Eric J. Bartelsman

ebartelsman@econ.vu.nl

Free University of Amsterdam and Statistics Netherlands

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

Abstract

This paper is the first of a series of explorations in the relative performance and sources of productivity growth of U.S. 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, 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: productivity, legal form of organization

* The authors would like to thank Carol Corrado for comments at various stages of the project and Jonathan Eller, Suzanne Polatz, Marcin Przybyla, Koen Vermelyen, and Matt Wilson for research assistance, and the BEA for providing data. This paper represents the authors own views and not those of Statistics Netherlands, the Board of Governors of the Federal Reserve System, or their staffs.

I. Introduction

In a self-referential application of the old dictum, "You don't often find $20 bills lying on the sidewalk", interesting questions and ready-made data sets are not found side-by-side. Especially for productivity analysis, where the subject of interest is the relation between outputs and inputs, researchers usually have to build their data sets using 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 gross output, labor input, capital services inputs, 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 their descriptions 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 available 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 and pieces of information. Second, the paper sketches the construction of a dataset to study productivity using this system. 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 U.S. is the focus of the paper and is prominent in the discussion of the data problems and in the examples given. However, the data issues are ubiquitous, and the system can accommodate other types 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 data sets 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 used to produce their estimates may be useful.[1] 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 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, the BLS and the BEA have different definitions of nonfarm business. The BLS 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.

[1] 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 (1993) 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.

A more fundamental problem is related to the underlying data collection. One well-known example of this is the firm-establishment problem. 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 (Postner, 1984). 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 lines of work of the individual establishments. Data collected at the firm level will classify all of GE in one industry based on its major line of work. 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) data set, economists in the Industry Division at the Bureau of Economic Analysis (BEA) have converted all of the data to an establishment-basis concept. The NIPAs, on the other hand, also present some industry data, but they are not consistent across different types of income. The compensation data (table 6.2C) are collected at the establishment level, and as table 1 illustrates, the two sources match. The profit data (table 6.16C), 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, personal, 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.

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 (table 1.14), indirect business taxes, business transfers, and the surplus of government enterprises less subsidies are presented separately, but for corporations (table 1.16) only the sum of the three 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 GPO data are not released until November. 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.

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, 1991). However, the hierarchical structure can be stored in related tables in the form of an ordered graph and can be traversed with recursion or 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 the tools as falling in one of four categories: aggregating, disaggregating, balancing, and concording 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:

Relational Database

Sector	Activity	Product	Transaction	Date	Units	Value
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:

Inputs Used by Industry

Date	**p241_03001**	**n80_emp**
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; some how the user knows that the units of employees are millions of workers, while intermediate goods are measured in billions of dollars.[2] 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 simply 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.

[2] 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. Within the NIPAs, data are reported at different levels. Industry data (section 6) are reported in millions of dollars,

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 a business 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 dummy business. The second imputed sector involves the rental value of non-profit's capital equipment and structures. As for owner-occupied housing, the BEA pretends there is a business that owns this capital and rents it to non-profit organizations.[3] The advantage of this approach is that it allows the BEA to treat the purchase of new owner-occupied housing by individuals and all capital equipment and structures owned by non-profits as investment. At the same time, it preserves the model of the circular flow whereby purchases of goods and services by businesses are either intermediate or investment, while the same purchases by households and non-profits that primarily serve individuals are consumption. 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.

By legal form of organization, the BEA includes these two imputed sectors within the

---

while data elsewhere are in billions of dollars.

[3]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 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.

category `Other Private Business', which is a subset of noncorporate business. We could account for these imputed sectors as simply one of three types of other private business, but we have found in some of our routines that it is easier to accomplish the same goal with a different dimension, rather than a refinement of the hierarchy in an existing dimension. It is useful to keep track of these imputed sectors because, as already discussed, the BLS measure of nonfarm output equals the BEA measure of nonfarm output less these two imputed sectors.

The dimensions that we use are summarized below:

Sector	represents the NIPA institutional sectors (business, government, households, non-profit institutions). The business sector is further refined by legal form of organization (corporate, noncorporate, etc.).
Activity	describes the particular industry that defines the producer, such as agriculture, manufacturing, etc. Sector and activity are not the same concept, but they are sometimes related. There are numerous classification schemes for industries, such as NAICS, NACE, ISIC, SIC, etc.
Imputed	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.
Transaction	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.
Product	represents the type of goods or services produced, purchased, consumed or supplied. As with activity 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.
Date	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, such as average over period, end of period, etc.

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.

Note that 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 (such as table 2.1 on Personal Income or all of table 3.1 on Government Receipts and Expenditures). Such an analysis would be necessary when studying income dynamics or general equilibrium. Although our current implementation of the system excludes the dimensions that describe the receiving entities, expanding the system to include additional dimensions later would be straightforward.

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, 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. For example, rounding issues make the subaggregates read directly from the data source more accurate than anything that can be calculated, at least 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 equals factor and nonfactor income is represented by the following transactions hierarchy:

- Value added
  - Non-factor income
    - Capital consumption
    - Taxes, transfers less subsidies
      - Indirect taxes
      - Business transfers
      - Subsidies (negative)
  - Factor Income
    - Compensation
    - Capital income
      - Profits
      - Proprietors' income
      - Rental income
      - Net interest

Suppose at the same time, there is another hierarchy that divides GDP between the value added of manufacturing and nonmanufacturing industries. Then, factor and nonfactor income 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.

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 splitting each between general government (GG) and government enterprises (GE). Other times the order is reversed.

GOV			
FED		SL	
GGF	GEF	GGSL	GESL

GOV			
GG		GE	
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 involves the treatment of government enterprises. Sometimes, government enterprises are counted as a business; other times they are treated as part of the government. This implies different aggregates, even when, as above, the bottom nodes of the two systems match one-for-one.

DOMESTIC TOTAL					
BUS			HH & INS		GG
CRP	NCPB	GE	HH	INS	

DOMESTIC TOTAL					
PRIV BUS		HH & INS		GOV	
CRP	NCPB	HH	INS	GG	GE

It is obvious from the above examples that the construction of hierarchies and the number of dimensions are not unique. There are different ways to slice the data. In the discussion on imputed values, we noted that instead of creating a separate dimension, one could simply refine the hierarchy and create three nodes underneath other private business. The issue also arose when considering different classification schemes. Another example was suggested by the two hierarchies that describe the government. Instead of having two hierarchies, one could create separate dimensions:

METHOD 1	**METHOD 2**	
SECTOR	**SECTOR 1**	**SECTOR 2**
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 data set 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 the defined hierarchy. Slipping subaggregates into the hierarchy meta data is all that is needed to calculate them easily. One can also apply the aggregation procedures to chain-aggregated data by providing auxiliary information, such as nominal weights. The procedure can handle weighted and unweighted averages and sums, as well as Laspeyres, Paasche, Fisher ideal, and Divisia aggregation.

II.E.2. Disaggregating

A second operation that is often required is disaggregation. 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 same types of operations for aggregation are accommodated by the disaggregation tool, such as sums, weighted averages, and Fisher-ideal indexes. The difference between aggregation and disaggregation, however, 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 the aggregate. On the other hand, disaggregation is a one-to-many operation, and thus, one needs additional information to determine which of the infinite possible ways to split the total. 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 a consistent estimate of the pieces.

For simple data that add, our procedure scales up or down the pattern data to yield disaggregated pieces that sum to the known total. 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. As with the aggregation procedure, one likely has a series of data that vary over other dimensions, and the disaggregation procedure can repeat this procedure for every value of the other dimensions.

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 gets the rough magnitudes right; in the limit, the fall back could be to simply split the aggregate evenly. 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, H that is to be split among two industries, h_1 and h_2. Using observed compensation, c_1, c_2 implicitly assumes that the hourly compensation rates in the two industries are the same. Using c to split H yields:

$$\frac{c_1}{c_1 + c_2}H, \quad \frac{c_2}{c_1 + c_2}H$$

$$\frac{w_1 h_1}{w_1 h_1 + w_2 h_2}H, \quad \frac{w_2 h_2}{w_1 h_1 + w_2 h_2}H$$

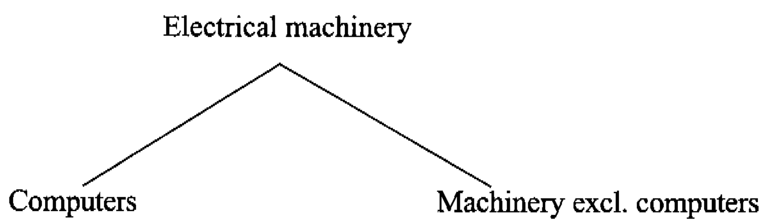
$$w_1 = w_2 = w \Rightarrow$$

$$\frac{wh_1}{wh_1 + wh_2}H, \quad \frac{wh_2}{wh_1 + wh_2}H$$

$$h_1,\ h_2$$

Industries that are aggregated together tend to be close in the classification scheme, and hence, similar in structure, opening up the possibility of employing this strategy 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 standard set of statistical assumptions, one could estimate standard errors around the estimates. By creating a simple hierarchy, one can calculate Fisher price indexes or chain aggregates that exclude a particular component. For instance, suppose one has data on total machinery manufacturers (industry 35) and computer and related equipment manufacturers (357). By writing the simple hierarchy, and applying the disaggregation tool, one could estimate the same data for machinery excluding computers. In this case, where only one disaggregated piece is constructed by 'subtraction', no pattern is necessary.



II.E.3. Balancing

A third operation balancing allows one to estimate data subject to linear constraints in multiple dimensions. Because the meta data contain all the information on the relevant constraints, our procedure is coded without explicit reference to the data structure of the problem at hand. 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 have been proposed in the literature.[4] 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 columns represent different values in one dimension, and the rows 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_1	T_2	$\dots$	Totals
Industries	I_1	a_{11}	a_{12}	$\dots$	$\sum_{j=1}^{J} a_{1j}$
	I_2	a_{21}	a_{22}	$\dots$	$\sum_{j=1}^{J} a_{2j}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	
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 sum 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 multiplies A_{k-1} by R_k so that $R_k A_{k-1}$ satisfies the column controls. Then one 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.[5] The limiting condition, $A = R A S$, also explains the moniker 'RAS' algorithm that 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 the algorithm is known as biproportional matrix balancing.

Schneider and Zenios (1990) credit Bregman for the result under a specific set of conditions that the RAS algorithm is also the solution to the problem of minimizing the cross entropy of A with the initial matrix A_0 subject to the row and column controls, a non-negativity constraint, and the biproportional constraint:

$$\begin{aligned}
&\min_{a_{ij}} \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\ln(a_{ij}/a_{ij}^{0}) \\
&\quad s.t. \sum_{i=1}^{m} a_{ij} = u_j \\
&\qquad \sum_{j=1}^{n} a_{ij} = v_i \\
&\qquad a_{ij} \geq 0 \\
&\qquad A = RA^{0}S \quad where\ A = [a_{ij}]
\end{aligned}$$

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 the error term ε_i has a known distribution. Two distributions that we allow for is the normal and log normal distribution, where the error term can be heteroskedastic but uncorrelated. 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 to also be measured with error.

Under the normal or lognormal assumptions, the weighted least squares solution is also a maximum likelihood problem:

$$\begin{aligned}
&\min_{a_i} \quad \sum_{i=1}^{N} \tfrac{1}{\sigma_i}(\log(a_i) - \log(a_i^0))^2 + \sum_{k=1}^{K} \tfrac{1}{\sigma_k}(v_k - \sum_{i=1}^{N} \phi_i^k a_i) \\
or \quad &\min_{a_i} \quad \sum_{i=1}^{N} \tfrac{1}{\sigma_i}(a_i - a_i^0)^2 + \sum_{k=1}^{K} \tfrac{1}{\sigma_k}(v_k - \sum_{i=1}^{N} \phi_i^k a_i)
\end{aligned}$$

If $\sigma_k = 0$, that is, the controls are measured exactly, then λ_k replaces $1/\sigma_k$ in the minimization problem where λ_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 the controls were measured exactly.

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 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.[6] 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.[7]

II.E.4. Concording

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:

SIC 1972	**SIC 1987**
Telephone, telegraph, and other communications services	Telephone, telegraph and other communications services
Radio & TV broadcasting	Radio & TV broadcasting, and other TV services
Wholesale trade	Wholesale trade
Retail trade	Retail trade

As is suggested by the above figure, the problem of concording from the left-hand to the right hand side 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 aggregating and disaggregating. 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. As a result, 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).

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 concordance weights as a balancing problem where the 1972 and 1987 totals are controls. As initial guesses for the pattern, 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:

Gross Output of Communications, 1987

			SIC 87	
			481,2,9	483-4
			157.8	42.1
SIC	481,2,9	170.1	157.8	12.3
72	483-4	29.7	0.0	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. Concording data from the 1972 SIC to the 1987 SIC and then back again yields the original 1972 data.

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

II.F. Specialized Productivity Tools

We have developed several tools to help in the study of productivity: one to calculate capital stocks by accumulating investment flows using a user-specified functional form for depreciation and discards; a second to calculate capital services from detailed estimates of capital stocks along with various models of user costs; and a third to compute and aggregate total factor productivity (TFP).

II.F.1. Estimating capital stocks

At its most basic level, the capital stock is modeled as the weighted sum of past investment using the so-called perpetual inventory method (PIM). Typically, the weights are less than or equal to one, reflecting the idea that equipment or structures bought in the past becomes increasingly less productive. This decay reflects both basic wear and tear, as well as an average rate at which firms scrap their old equipment. Let t i ndex time, s the vintage or the time the investment was made, j the type of equipment or structure, and it he activity and sector that holds the capital. Then the total productive stock of each asset type equals:

$$K_t^{i,j} = \sum_{v=0}^{\infty} f(I_{t-v}^{i,j}, \theta_{t,t-v}^{j}, j, s),$$

where θ is a set of parameters that presumably depends only on the relative age of the vintage and type of equipment. Only for some asset types is there information on the variability of θ over activities; information on θ changing over time is available for quite a few asset types. Various specifications have been proposed for the function f; we currently have coded the standard BLS method with a stochastic mean service life and beta-decay, as implemented at the FRB by Gilbert and Mohr (1996). We also have coded the function used by the BEA to calculate wealth stocks in order to split capital consumption allowances between the consumption of fixed capital and the capital consumption adjustment (Fraumeni, 1997).

The procedure that we have coded treats the parameter vector and the level of investment as data, linked and organized by meta data. This feature allows researchers much flexibility in changing the various assumptions needed for the PIM.

In the procedure, an initial capital stock is estimated based on real investment growth in the first ten years for each particular unique activity-by-asset-type combination and on its steady state geometric depreciation rate. The steady-state depreciation rate is calculated by finding the productive decay of a steady-state stock resulting from a constant investment rate run through the PIM system for each of the parameter combinations used by the BLS. This geometric rate is also an output because it is used in the computation of the user cost of capital.

In our application, with real investment split by GPO activity and by legal form of organization, BEA asset types, and time, we need to manipulate about 140,000 data points. When the data are interacted with the full vintage structure and decay function, about 7 million records are processed. Computation of the full set of capital stock time series takes less than 15 minutes on a 700Mhz Pentium III notebook, with a relatively slow 4200rpm hard disk.

II.F.2. Estimating capital services

Capital services for a particular (sub)set of asset types can be computed as a Divisia aggregate of the underlying productive stocks, using implicit expenditures on each asset type as weights. Expenditure flows for each asset are computed by multiplying the stock with a user cost. The user cost procedure does not depend on the exact structure of the underlying data, but runs off the supplied data and meta-data. The procedure needs a dataset that varies over time for dimensions other than asset type (in our case, by activity and LFO) containing estimates of property type-income and other necessary series that do not vary by asset type. A second input dataset provides data that vary over all relevant dimensions and contains real capital stocks, depreciation rates, a capital gains term, *dq*, and relevant tax parameters that vary by asset (and possibly by activity). The model of user costs that we employ is the standard formula (Hall and Jorgenson, 1967). Let ρ_i^j denote the user cost of asset *j*, varying over activity *i*(t ime subscripts omitted) and PTI_i equal the property-type income of activity *i*. Then:

$$\rho_i^j = (r + d_{ij} - dq_{ij}) * \frac{(1 - ITC_{ij} - pdvZ_{ij})}{1-\tau}$$

$$PTI_i = \sum_j P_{ij} * K_{ij}$$

where δ is the geometric depreciation consistent with the PIM method used, *dq* is a capital gains term, *ITC* is the investment tax credit, *pdvZ* is the present discounted value of future tax deductions for depreciation, τ is the corporate tax rate, and *P* is the asset price deflator. Users supply all the necessary data to compute the rate of return, *r*, for each industry and the user cost ρ. A sub-procedure is built in to calculate the capital gains term, *dq*, using a two period moving average of changes in the asset deflator. Instead of calculating the ex-post rate of return, the system can compute the user cost based on a supplied estimate of the (ex-ante) rate of return, *r*.

II.F.3. Estimating TFP

We have also built a simple tool to allow for the different schemes employed to weight the rates of change of the factor inputs used to calculate TFP. Following the choices made in the literature, the user determines whether the weights should be cost- or revenue shares, and with revenue shares, the user can choose whether or not to impose constant returns to scale. Alternatively, the user may supply factor shares for each input (e.g. from regression output, or from other countries to do cross-country comparisons of TFP levels). The procedure calculates TFP growth as well as factor shares, output contributions, and rates of change for all inputs for each combination of the given dimensions.

To calculate aggregate TFP growth, the user supplies a set of weights, which need to be consistent with the inputs and output concepts used. If a user were to model gross output as a function of intermediate inputs and other factors of production, Domar weights would be a natural choice. Alternatively, if disaggregated TFP were computed for value added, then value-added weights would be appropriate.

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 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, indirect business taxes, 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.[8] Data on employment and all persons engaged in production are also included in the dataset. Complete data are available from 1987 to 2000, 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; the procedure to concord the data was explained above. Second, we filled in some missing data. Table 2 describes which data had to be estimated and the method by which the estimates were made. The table is broken into sections. The particular data that was estimated is listed at the top of each section in bold. The first two columns report the particular industries and years that apply. For some estimates, the BEA provided to us unpublished data. Otherwise, typically the data were estimated by disaggregating an observation or by estimating the data subject to some restrictions. The sources for these controls, as well as the initial estimates of the pattern of the detailed data, are listed in the third column. Other details of the disaggregation or balancing tools are described in the last column.

A few additional comments about some of the estimates are in order. 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 gross product deflators, these yielded gross output deflators.

The GPO 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. It is 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

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. We also allocated the imputations (owner-occupied housing and the rental value of the capital of non-profits) across industries. Table 3 describes the sources and methods by which we made these estimates; the table is organized in the same manner as table 2.[9]

Refining these data by legal form of organization is useful for several reasons. Splitting owner-occupied housing and the rental value of non-profits' capital allows us to better account for the productivity of the BLS's nonfarm corporate sector.[10] In the case of owner-occupied housing, no labor input is associated with this output. In the case of non-profits, their labor input and labor compensation are counted outside of business. The BLS adjusts the BEA's nonfarm output to exclude these imputations when estimating productivity because, as constructed, they represent output generated without any labor input. One may also want to examine these data separately because the pricing, hiring, and investment decisions of these organizations are presumably not motivated by profit maximization.

Splitting the industry output by legal form is also 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 data also have to be 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 to the corporate sector.

Corrado and Slifman (1999) showed that productivity in the noncorporate business sector is 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 asset type that the BEA has made available on its web site (www.bea.gov/bea/dn/faweb/). Currently, these data are available through 2000, although we have concorded the data to the NIPA data on domestic investment by asset type in order to update them through 2001. We have refined these data by splitting industry investment between corporate and noncorporate investment for each type of equipment and structure. As also noted in table 3, we estimated these splits, controlling the total for each legal form to equal the data available in tables 4.7 of 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). As an initial pattern, we used property-type income, although in cases where this was negative, or too volatile, compensation adjusted for 'labor income' of proprietors was used as a pattern. 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.

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 TFP in the usual manner; TFP growth equals the growth rate of real gross output less the share-weighted growth rates of real intermediate inputs, labor input, and capital services. Our estimates of real output, intermediate inputs, and cost-weighted expenditure shares come from our modified GPO dataset. Labor input is measured as hours of all persons. Capital stocks are calculated by accumulating the investment data using the standard BLS stochastic mean-service life and beta-decay parameters, and from that, capital services are estimated using ex-post returns.

Tables 4a-4c report these estimates for the fifty-nine industries that we analyze over three selected time periods. The columns at the far right list the annual growth rate of real output. This equals the sum of the annual growth rate of total factor productivity plus the contributions from labor, three types of capital, and materials. We report separate estimates of the contribution of high-tech capital (computers, software, and communications equipment), other capital equipment, and structures.

To calculate aggregate TFP growth we take a weighted sum of the individual components, where the weights are calculated as sketched in Domar (1961).[11] 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, agriculture services, 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.[12] The left-hand columns of tables 4a-4C report the average of the Domar weights in each sub-period for each industry. The weighted sums of TFP growth 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 5 reports aggregate TFP for private businesses excluding agriculture over each of the time periods considered, as well as an estimate of the contribution to growth from increases to the stock of high-tech capital equipment. We estimate that TFP accelerated 0.8 percentage point to 1.2 percent per year, on average, in the 1996-2000 period. These results are similar to the BLS's measure of multifactor productivity.

The rest of table 5 repeats the information in tables 4a-4c for some selected industries to illustrate how we can trace the sources of the pickup in TFP. As has been noted elsewhere, although these technical progress has always been rapid in the machinery manufacturing (which includes computers) and electrical machinery manufacturing (which includes communication equipment and semiconductors) industries, these industries contributed importantly to the acceleration in TFP. Nonetheless, total factor productivity growth shot up in the last half of the 1990s. 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; total factor productivity growth, which measures what is left over after accounting for from the acceleration in capital services, did not increase in these industries, however.

Table 6 reports total factor productivity growth split between corporate and non-corporate private businesses. At the aggregate level, total factor productivity accelerated in both forms of legal organization, but in each period, corporate productivity grew faster than for other types of businesses. This last observation, however, appears to be a function of the mix in the legal form of business across industries that differ in their overall rates of technical advance. Examining a set of two-digit service industries that have both types of businesses shows no coherent pattern. Only in the health industry did the productivity of corporations increase more rapidly in each sub-period. For other industries, such as auto repair, services, and parking, noncorporate businesses appear to have outperformed corporations. A closer examination of this issue at an even more disaggregated level may be warranted.

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.

The issue for the study of productivity is that these stories suggest that the 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 the rate of depreciation and discards. Unless augmented, they assume that this 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 was spending a little over $8 billion and a Federal Reserve study that suggests 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 these data to estimate baseline spending 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 goosed up the fraction on software by 50 percent based on some IDC figures on the split on spending between hardware and software.

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 it 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 scrappage of older stocks of hardware and software.

Table 7 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. The largest effect on any aggregate in any year is +0.32 percent in the communications industry in 1997. The extra scrappage 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 0.32 percentage point. In a few industries, such as communications, depository and nondepository institutions, and 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. Assuming a more moderate level of Y2K spending that represents replacement investment (\$50 billion) reduces the cumulative effect to one-third 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 perform various manipulations 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, 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. In the aggregate, total factor productivity of corporations grew faster than of sole proprietorships, partnerships, and other private business. An examination of most service industries, however, suggests that a good deal of this result owes to the fact that the share of output from noncorporate businesses is highest in those industries with lagging productivity growth, regardless of the legal organization of firms.

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.

By no means is the productivity data set that we have constructed finished. Several of the assumptions that we made to create some estimates need to be reconsidered, especially those relating to price deflators. Data are available to further refine the noncorporate estimates by splitting them between sole proprietorships and partnerships versus other private business, and some data are available to allocate government enterprises across industries.[13] Adding in capital service estimates of government enterprises, the general government, private households, owner-occupied housing, and non-profits also will enable us to do a complete accounting of U.S. productivity, such as has been the general practice of Professor Jorgenson and his colleagues.

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, 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 start 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. 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

Bacharach, Michael (1965), "Estimating Nonegative Matrices from Marginal Data", *International Economic Review* 6(3), (September) 294-310.

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).

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.

Günlük-Senesen, 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.

Larkins, Daniel, (1992), "Business Situation", 72(9) (September), 1-4.

Larkins, Daniel, Ralph W. Morris, Eugene P. Seskin, and Frederick von Batchelder (2001), "Business Situation", 81(11) (November), 1-7.

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).

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.

Parket, Robert and Bruce Grimm (2000), "Recognition of Business and Government Expenditures for Software as Investment: Methodology and Quantitative Impacts, 1959-98", http://www.bea.gov/bea/papers/software.pdf.

Postner, Harry H. (1984), "New Developments Towards Resolving the Company-Estatblishment 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.

Parker, Robert P. and Eugene P. Seskin (1997), "The Statistical Discrepancy", *Survey of Current Business* 77(8) (August), 19.

Seskin, Eugene P. and Robert P. Parketr (1998), "A Guide to the NIPA's", *Survey of Current Business* 78(3) (March), 26-68.

Sinkhorn, R. (1964), "A Relationship between Arbitrary Positive Matricies 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.

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.

U.S. Department of Commerce, Economics and Statistics Administration (1999), "The Economics of Y2K and the Impact on the United States", November 17.

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 2000 Compensation and Profits Data in the GPO and the NIPA Datasets*

	Compensation		Profits w/ IVA	
	GPO	NIPA	GPO	NIPA
Manufacturing	979.4	979.4	143.3	155.2
Transportation and Utilities	374.3	374.3	63.4	67.4
Wholesale Trade	386.6	385.6	58.0	60.5
Retail Trade	510.4	510.4	78.8	81.8
Remaining domestic private industries	2461.8	2461.8	352.8	331.4

*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
Estimating Industry Data

Industries	Year	Controls and initial guess	Method
Split full-time equivalent employees (FTE), full-time and part-time employees (FTPT), and persons engaged in production (PEP) of 65 (SIC 87) and 65A6 (SIC 72) between 65re and 65hs.			
65re and 65hs	1950-2000	□ Sum of 65re and 65hs from GPO. □ Use compensation as initial guess.	Disaggregation used.
Estimate employee hours by industry			
All private	1987-1997	Direct from table BEA.	
All private	1998-1999	□ Totals of available subaggregates from table 6.9c. □ Initial pattern from old BEA data.	Disaggregation used. BEA provided hours data that have been subsequently revised.
All private	2000	□ Totals of available subaggregates from table 6.9c. □ Use compensation as initial pattern.	Disaggregation used.
All private ex. 65re, 65hs, 83, 86	1971 1986	Direct from BEA.	
83, 86	1971-1974	□ Hours for 83 plus 86 calculated as hours for all services less hours in other service industries. □ Use compensation as initial pattern.	Disaggregation used.
All private, incl. 65, ex. 65re, 65hs, 83, 86	1949-1970	□ Totals of available subaggregates from table 6.9b. □ Initial pattern calculated in steps: □ Regress log(HOURS) on log(FTE), log(COMP), log(COMP/Hours)) for the available aggregate, and log(HOURS/FTE) for the aggregate 1971-1986. □ Use estimated coefficients to calculate E[log(HOURS)] over period. □ Use E[HOURS] = exp(E[log(HOURS) + ½ variance)]) as initial pattern.	Disaggregation used.
83, 86	1949-1970	□ Same as above except exclude log(FTE) and log(HOURS/FTE) from regression.	Disaggregation used.
65re, 65hs	1949-1986	□ Total for 65+66 calculated above or in GPO. □ Use compensation as initial pattern.	Disaggregation used.

Table 2, continued

Industries	Year	Controls and initial guess	Method
Estimate non-employee hours by industry.			
Farm	1959-2000	Direct from BLS: Business all persons hours - nonfarm busines all persons hours minus the same for employee hours.	
Nonfarm private	1959-2000	□□ Manufacturing total from BLS (All persons hours less employee hours) □□ Nonmanufacturing total from BLS (All persons hours, nonfarm less employee hours, nonfarm) less manufacturing total. □□ Initial pattern equals proprietors' income times compensation per employee hours.	□□ Disaggregation used. All non-employee hours are non-corporate. Corporate non-employee hours equals zero.
Estimate self-employed compensation by industry.			
Farm	1959-2000	Direct from BLS (all persons compensation less employee compensation for business minus the same for nonfarm business).	
Non-farm	1959-2000	□□ Totals for durable manufacturing, nondurable manufacturing, and other non-farm direct from BLS (all persons compensation less employee compensation). Other non-farm calculated as nonfarm business less manufacturing. □□ Initial pattern equals self-employed hours times compensation per hour of employees.	Disaggregation used.
Estimate missing gross output and intermediate inputs.			
44, 47, 60, 61, 65re, 67, 73, 83, 84,9, 86, gov. ent.	1977-1986	□□ Concord 1987 input-output table to 1972 SIC. Take average of output to value added in 1987 concorded table and 1982 input-output table. Take 1982 ratio for 1982 and years before. Interpolate between 1982 and 1987 ratio for 1983-1986. Multiply ratio by gross product. □□ Intermediates equal gross output less gross product.	
36, 38	1977-1986	□□ Like procedure above, but use results as a pattern □□ Aggregate of 36 and 38 from GPO. □□ Intermediate inputs equals gross output less gross product.	Disaggregation used.

Table 2, continued

Industries	Year	Controls and initial guess	Method
Estimate missing gross product deflators.			
44, 47, 60, 61, 65re, 67, 73, 83, 84,9, 86, gov. ent.	1977-1986	Use concorded use tables calculated above to weight intermediate input prices. Use the gross output deflators of industries that are known as intermediate prices. Add in value added of industries to be estimated and the value added deflator in GPO. Use a guess for the gross output deflator for the missing industries to use as input prices for other industries. Iterate until convergence. Use gross product deflator as initial guess. Intermediate input deflators equal gross output deflator less (in chain weighted sense) gross product deflator.	
36, 38	1977-1986	Like procedure above, but use results as a pattern Aggregate of 36 and 38 from GPO. Intermediate input deflators equal gross output deflator less (in chain weighted sense) gross product deflator.	Disaggregation used.

Table 3
Estimating Industry-by-Sector Data

Industries	Year	Controls and initial guess	Method
Split compensation by industry and by sector.			
All private	1987-1997	Direct from BEA. BEA provided to us compensation by LFO for 1987-1999, but the 1998-1999 data have been revised.	
All private	1998-1999	Total for each industry from GPO. Total for each sector from table 1.15 Initial guess is the unrevised industry-by-sector data from BEA.	Balancing (log spec.) used. The NIPA data are adjusted for rounding to sum exactly to the GPO data. Standard deviations for all controls then set to zero. Standard deviations for guesses where comp = 0 is 0, otherwise = 0.1.
All private	2000	Controls as above. Initial guess equals 1999 data by industry-by-sector times the ratio of industry compensation in 2000 to industry compensation in 1999.	As for all private compensation (1998-1999).
All private excl. 83, 86	1959-1986	Direct from BEA Data are on 1972 SIC. Industry-by-sector data concorded to 1987 SIC assuming that for each industry, splits across sectors are the same as for the aggregate.	
83, 86	1975-1986	Direct from BEA.	
83, 86	1959-1974	Total for each industry from GPO. Total for each sector equals values from table 1.15 less sum of other industries. Initial guess for year *t* equals mix in year *t+1*.	Balancing (log spec.) used. The estimates are run iteratively backward starting in 1986. NIPA aggregates are corrected and standard deviations are calculated as in all private compensation (1998-1999).
Allocate all income of owner-occupied housing across industries			
All private	1959-2000	Direct from table 8.21. Only in industry 65, for other equals zero.	

Table 3, continued

Industries	Year	Controls and initial guess	Method
Allocate all income of rental value of non-profit's equipment and structures across industries.			
All private	1959-2000	National totals from table 8.21. Initial pattern equals compensation of institutions times ratio of each income in GPO divided by compensation in GPO.	Disaggregation used.
Allocate business transfer payments by industry and by sector.			
All private	1959-2000	Direct from GPO because all transfers are assumed to be corporate.	
Calculate non-corporate income excluding owner-occupied housing and rental value of non-profit's equipment and structures across industries.			
All private	1959-2000	Compensation for noncorporates calculated above; compensation for imputed sectors equals zero. Other components for noncorporates from GPO.	Direct subtraction.
Allocate indirect taxes plus subsidies by industry and by sector.			
All private	1987-1997	Direct from BEA. BEA provided to us the sum of indirect taxes, business transfers and subsidies for corporations for 1987-1999. The 1998-1999 data have been revised. All business transfers subtracted out to get indirect taxes plus subsidies.	
All private	1998-1999	National totals for corporate taxes plus subsidies sum to table 1.16 less business transfers, table 1.14 National totals for noncorporate taxes plus subsidies sum to NIPA total (table 1.14) less corporate total (table 1.16) and less government enterpirses in the GPO. Industry toals from GPO. Initial guess equals unrevised BEA data.	Disaggregation used.
All private	2000	Initial pattern equals the 1999 value times the ratio of value added less indirect taxes and subsidies in 2000 to the value added less indirect taxes and subsidies in 1999.	Disaggregation used.

Table 3, continued

Industries	Year	Controls and initial guess	Method
All private	1959-1986	▯ National and industry totals from GPO. ▯ The initial guess I_t equals the value in year $t+1$ times the ratio of value added less taxes and subsidies in year t to value added less taxes and subsidies in year $t+1$. The value for $t+1$ = 1987 is from above.	Balancing (log spec.) used. The estimates are run iteratively backward starting in 1986. NIPA aggregates are corrected and standard deviations are calculated as in all private compensation (1998-1999).
For each industry, split the aggregate of indirect business taxes less subsidies between taxes and subsidies and by sector.			
All private	1959-2000	▯ Industry by sector sum of taxes and subsidies calculated above. ▯ Industry total for taxes from GPO. ▯ Industry total for subsidies from GPO. ▯ Initial pattern for indirect taxes equals value added except indirect taxes, transfers and subsidies by industry and by sector times indirect taxes for whole industry divided by valued added less taxes, transfers and subsidies for whole industry.	For industries where subsidies equal zero, assume subsidies equal zero for all sectors; taxes equal the previously called total of taxes and subsidies. For industries where subsidies not equal zero, use Balancing (log spec.). For negative initial guesses, multiply guess and all coefficients in linear restrictions by -1. Treat owner-occupied housing and rental value of non-profit's capital as separate sectors.
Split corporate compensation into financial & non-financial			
All private excl. 67	1959-2000	▯ Industries 60-63 are financial. ▯ Industries excluding 60-63, 67 are nonfinancial.	
Industry 67	1959-2000	▯ Nonfinancial equals all corporations from table 1.16 less sum of other nonfinancial corporations calculated above. ▯ Financial equals corporate total for 67 less nonfinancial 67.	Calculated directly.

Table 3, continued

Industries	Year	Controls and initial guess	Method
Split employee hours by industry and by sector.			
All private		□ Industry total previously calculated. □ Nonfinancial corporate total from BLS. □ Initial pattern equals compensation by sector; imputed parts of non-corporate equals zero.	BALANCING (log spec.) used. Standard deviations for controls equal zero; other standard deviations equal 0.1.
Split all employees and full-time equivalent employees by industry and by sector.			
All private	1959-2000	□ Industry total from GPO. □ Initial pattern equals employee hours split by industry and by sector calculated above.	Disaggregation used.
Split non-employee hours by industry.			
Farm	1959-2000	Direct from BLS: business all persons hours - nonfarm business all persons hours minus the same for employee hours.	
Nonfarm private	1959-2000	□ Manufacturing total from BLS (all persons hours less employee hours). □ Nonmanufacturing total from BLS (all persons hours, nonfarm less employee hours, nonfarm) less manufacturing total. □ Initial pattern equals proprietors' income times compensation per employee hours.	□ Disaggregation used. All non-employee hours are non-corporate. Corporate non-employee hours equal zero.

Table 3, continued

Industries	Year	Controls and initial guess	Method
Split self-employed compensation by industry.			
Farm	1959-2000	Direct from BLS (all persons compensation less employee compensation for business minus the same for nonfarm business).	
Non-farm	1959-2000	Totals for durable manufacturing, nondurable manufacturing, and other non-farm direct from BLS (all persons compensation less employee compensation). Other non-farm calculated as nonfarm business less manufacturing. Initial pattern equals self-employed hours times compensation per hour of employees.	Disaggregation used.
Estimate legal-form splits of investment by industry and by asset type.			
All private	1901-2000	Total for each industry x asset type from unpublished BEA estimates available at http://www.bea.gov/bea/dn/faweb/. Total investment by legal-form for equipment and for structures from table 4.7 of Standard Fixed Asset Tables published by BEA. Use property-type income (value added less indirect taxes and compensation) or adjusted-compensation as a pattern.	Balancing (log spec.) used in each year.

Table 4a
Growth Accounting
(1977-1989)

		Contributions to Output Growth						
			Capital Services					
	Domar weight	Mat.	ICT	Oth. Eqp.	Str.	Lab.	TFP	Output
Metal mining	0.3	1.29	0.01	-0.28	0.17	-0.55	3.08	3.71
Coal mining	0.9	1.32	0.00	-0.13	0.19	-0.96	2.42	2.84
Oil and gas extraction	5.3	-0.80	0.19	0.15	0.68	-0.18	-0.59	-0.54
Other mineral mining	0.4	-0.45	0.03	-0.07	0.18	0.05	0.71	0.45
Construction	12.7	0.23	0.01	0.00	0.04	1.11	-0.48	0.90
Lumber and wood	1.6	0.43	0.15	0.09	0.03	0.10	0.38	1.19
Furniture & fixtures	1.1	1.59	0.05	0.11	0.09	0.38	0.28	2.50
Stone, clay, and glass	1.8	0.31	0.04	-0.02	-0.02	-0.15	0.30	0.45
Primary metals	4.0	-0.59	0.02	0.00	0.06	-0.79	-0.20	-1.49
Fabricated metals	4.6	-0.22	0.02	-0.01	0.03	-0.10	0.58	0.30
Machinery	6.5	1.80	0.02	0.06	0.11	0.11	2.82	4.93
Electrical machinery	4.4	2.27	0.09	0.20	0.05	0.44	2.66	5.71
Motor vehicles	5.0	1.10	0.20	0.05	0.09	-0.24	-0.59	0.62
Other transportation equip.	3.2	2.34	0.09	0.05	0.09	0.90	0.38	3.85
Instruments	3.4	2.51	0.05	0.03	0.01	0.50	0.98	4.06
Miscellaneous mfg.	1.1	-0.83	0.12	0.11	0.06	-0.20	0.93	0.19
Food	9.1	1.36	0.01	-0.05	0.00	0.02	0.80	2.13
Tobacco	0.6	3.45	0.45	-0.21	0.00	-0.31	-4.53	-1.14
Textiles	1.5	-0.53	0.06	-0.08	-0.03	-0.35	0.88	-0.05
Apparel	1.9	-0.15	0.13	0.00	0.01	-0.36	1.03	0.66
Paper	2.6	1.78	0.47	0.05	0.03	0.04	-0.08	2.29
Printing	3.4	2.35	0.44	0.17	0.09	0.91	-0.77	3.20
Chemicals	5.7	1.08	0.07	0.00	0.07	0.04	0.59	1.84
Petroleum refining	6.0	-1.06	0.19	0.05	0.06	-0.09	0.48	-0.36
Rubber and plastics	2.5	1.71	0.28	0.03	0.09	0.40	1.08	3.58
Leather	0.3	-3.32	0.11	-0.06	0.02	-1.59	0.88	-3.95
Railroad transportation	1.1	0.80	0.06	-0.19	-0.10	-2.32	3.85	2.08
Local & interurban transit	0.5	-0.65	0.07	0.01	0.13	0.96	-1.45	-0.93
Trucking & warehousing	3.3	2.50	0.02	0.17	0.01	0.44	0.42	3.55
Water transportation	1.1	-4.48	0.01	-0.08	0.06	0.10	0.51	-3.88
Air transportation	1.9	3.35	0.11	0.08	0.10	2.36	0.43	6.43
Pipelines ex. natural gas	0.3	1.24	0.07	-0.07	-0.21	-0.09	-1.16	-0.22
Transport services	0.4	1.97	0.23	0.09	0.04	3.08	-0.05	5.36
Telephone & telegraph	4.0	2.22	0.83	0.10	0.63	0.07	1.60	5.45
Radio and television	1.1	2.56	0.43	0.07	0.65	0.62	-2.14	2.19
Utilities	6.8	0.59	0.30	0.24	0.27	0.32	-1.10	0.62
Wholesale trade	14.0	0.50	0.26	0.10	0.28	1.26	1.27	3.67
Retail trade	18.5	0.92	0.22	0.05	0.32	1.47	0.01	2.98
Depository institutions	4.2	2.96	1.18	0.66	1.14	1.11	0.15	7.20
Nondepository institutions	.7	2.84	2.29	1.77	0.21	1.82	0.36	9.29

Table 4a, continued
Growth Accounting
(1977-1989)

	Domar weight	Contributions to Output Growth: Mat.	Capital Services: ICT	Capital Services: Oth. Eqp.	Capital Services: Str.	Lab.	TFP	Output
Security/commodity brokers	1.2	3.37	0.48	0.13	0.43	3.32	1.55	9.28
Insurance carriers	3.5	3.55	0.46	0.18	0.54	0.67	-3.00	2.40
Insurance agents & brokers	1.3	0.87	0.15	0.00	0.16	1.37	-0.25	2.29
Nonfarm housing	3.5	0.04	0.00	0.00	0.05	0.03	0.00	0.29
Other real estate	7.4	0.33	0.05	0.09	0.05	0.31	-0.19	0.64
Investment offices	0.5	6.32	1.22	0.65	1.21	2.51	3.60	15.52
Hotels	1.7	1.33	0.04	0.03	0.14	1.31	-1.27	1.58
Personal services	1.4	0.89	0.03	0.02	0.04	1.00	-0.32	1.65
Business services	4.4	2.50	1.28	-0.12	0.08	3.81	-0.34	7.22
Automobile services	1.9	1.10	0.09	0.43	0.08	1.11	-0.28	2.53
Misc. repair services	0.8	1.63	0.08	0.02	0.05	1.06	0.14	2.97
Motion pictures	0.5	2.21	0.06	0.18	0.11	0.83	0.21	3.59
Amusement & recreation	1.0	1.80	0.00	-0.08	0.08	0.85	0.71	3.35
Health services	6.3	2.77	0.21	0.07	0.27	1.53	-1.00	3.85
Legal services	1.9	0.78	0.20	0.19	0.20	1.92	-0.67	2.62
Educational services	0.7	3.41	0.04	0.02	0.18	-0.01	-0.26	3.39
Social services	0.6	7.96	0.11	0.03	0.02	1.49	-0.74	9.19
Membership organizations	0.7	2.58	0.10	0.01	0.03	0.58	-0.10	3.21
Other services	3.7	2.38	0.27	0.03	0.06	2.03	0.70	5.47

Table 4b
Growth Accounting
(1990-1995)

		Contributions to Output Growth						
			Capital Services					
	Domar weight	Mat.	ICT	Oth. Eqp.	Str.	Lab.	TFP	Output
Metal mining	0.2	-0.05	0.34	0.03	-0.01	-0.35	2.74	2.71
Coal mining	0.5	-1.03	0.10	-0.02	0.07	-1.06	2.48	0.54
Oil and gas extraction	2.0	-0.39	0.10	-0.29	0.09	-0.33	0.40	-0.42
Other mineral mining	0.3	-0.07	0.15	0.08	-0.01	-0.07	0.61	0.69
Construction	9.6	-0.15	0.07	0.04	0.03	-0.07	-0.09	-0.17
Lumber and wood	1.3	1.31	0.18	0.18	0.02	0.01	-1.25	0.44
Furniture & fixtures	0.9	1.36	0.03	-0.08	0.00	-0.15	0.68	1.84
Stone, clay, and glass	1.2	-0.30	0.12	-0.02	-0.01	-0.25	0.89	0.43
Primary metals	2.5	0.71	0.05	0.00	0.02	-0.30	0.78	1.28
Fabricated metals	3.4	0.87	0.05	-0.05	0.02	0.03	0.93	1.84
Machinery	5.2	4.46	0.04	0.04	0.03	-0.09	1.74	6.22
Electrical machinery	4.2	4.68	0.12	0.23	0.06	-0.39	5.65	10.34
Motor vehicles	4.5	1.90	0.17	-0.03	0.03	0.59	0.74	3.41
Other transportation equip.	2.5	-1.69	0.27	0.05	0.14	-1.94	-1.43	-4.60
Instruments	2.7	2.69	0.08	0.07	-0.03	-1.05	-0.45	1.32
Miscellaneous mfg.	0.8	1.85	0.26	0.41	0.09	0.06	-0.39	2.29
Food	7.3	0.71	0.03	-0.09	-0.05	0.09	0.92	1.61
Tobacco	0.7	1.04	0.06	-0.09	-0.03	-0.32	-0.18	0.48
Textiles	1.1	0.72	0.05	-0.03	-0.03	-0.39	1.33	1.65
Apparel	1.4	1.19	0.20	-0.01	0.01	-0.66	0.36	1.09
Paper	2.4	1.49	0.39	-0.02	0.02	0.02	-0.35	1.55
Printing	3.2	0.86	0.31	0.17	0.04	-0.07	-1.94	-0.62
Chemicals	5.2	-0.39	0.11	0.20	0.08	-0.06	0.60	0.53
Petroleum refining	3.0	0.48	-0.01	-0.05	0.06	-0.13	-0.02	0.34
Rubber and plastics	2.4	2.83	0.33	-0.03	0.03	0.44	0.94	4.53
Leather	0.2	-2.51	0.36	0.03	0.05	-1.43	1.07	-2.43
Railroad transportation	0.7	0.34	0.11	-0.04	-0.11	-0.83	3.66	3.13
Local & interurban transit	0.4	0.85	0.01	0.09	0.14	1.16	-1.36	0.89
Trucking & warehousing	3.1	2.41	0.11	0.11	0.03	0.20	1.07	3.92
Water transportation	0.5	1.23	0.01	-0.15	0.01	0.29	0.98	2.36
Air transportation	1.9	-0.71	0.17	0.06	0.01	3.17	-0.04	2.65
Pipelines ex. natural gas	0.2	1.78	0.77	0.04	-0.13	-0.33	-3.51	-1.40
Transport services	0.6	2.22	0.80	-0.12	0.09	1.87	-0.43	4.43
Telephone & telegraph	4.2	2.31	0.57	0.24	0.21	0.05	1.30	4.68
Radio and television	1.1	-2.38	1.07	0.11	0.68	0.41	1.32	1.20
Utilities	6.2	-0.01	0.20	0.25	0.29	-0.04	0.14	0.83
Wholesale trade	13.0	2.09	0.16	0.17	0.26	0.10	1.00	3.77
Retail trade	18.1	1.36	0.42	0.08	0.20	0.45	0.03	2.53
Depository institutions	5.5	0.91	0.71	-0.43	0.66	-0.53	0.00	1.33
Nondepository institutions	1.2	5.05	2.27	0.66	0.08	1.24	-1.60	7.69

Table 4b
Growth Accounting
(1990-1995)

		Contributions to Output Growth						
			Capital Services					
	Domar weight	Mat.	ICT	Oth. Eqp.	Str.	Lab.	TFP	Output
Security/commodity brokers	2.0	3.46	0.06	0.20	0.19	1.65	2.18	7.73
Insurance carriers	4.4	-0.62	0.56	0.15	0.22	0.25	-0.34	0.22
Insurance agents & brokers	1.3	0.49	0.25	0.24	0.17	0.38	-2.51	-1.00
Nonfarm housing	3.6	-0.03	0.00	0.00	0.03	0.29	0.00	0.16
Other real estate	8.5	0.61	0.20	0.04	0.07	0.03	0.19	1.15
Investment offices	0.4	1.16	-0.05	-0.02	0.91	0.35	1.87	4.23
Hotels	1.8	0.84	0.05	0.05	0.05	0.19	0.73	1.90
Personal services	1.5	1.03	0.13	0.01	0.08	0.22	-0.27	1.21
Business services	7.0	2.73	0.34	0.20	0.04	2.11	0.85	6.28
Automobile services	2.0	1.26	0.04	1.13	0.07	0.42	-0.84	2.07
Misc. repair services	0.8	2.32	0.20	0.02	0.06	-0.06	-1.15	1.38
Motion pictures	0.6	0.96	0.18	0.08	0.12	0.53	-1.56	0.31
Amusement & recreation	1.2	3.40	0.03	0.13	0.15	1.36	-0.22	4.85
Health services	8.2	2.25	0.19	0.02	0.17	1.14	-1.02	2.74
Legal services	2.4	0.18	0.07	-0.03	0.03	0.28	-0.05	0.47
Educational services	1.8	2.67	0.03	-0.03	0.10	0.39	-0.48	2.69
Social services	0.9	2.95	0.11	0.01	0.05	1.18	-0.62	3.67
Membership organizations	0.8	3.17	0.05	-0.01	0.02	-0.18	0.68	3.73
Other services	4.8	1.31	0.15	-0.02	0.02	1.18	-0.03	2.61

Table 4c
Growth Accounting
(1996-2000)

	Domar weight	Contributions to Output Growth: Mat.	Capital Services: ICT	Capital Services: Oth. Eqp.	Capital Services: Str.	Lab.	TFP	Output
Metal mining	0.1	-2.87	0.11	-0.05	-0.09	-1.46	4.59	0.23
Coal mining	0.4	-2.47	0.11	0.17	0.06	-1.11	1.04	-2.21
Oil and gas extraction	1.8	3.48	0.33	0.09	0.26	-0.32	-3.49	0.35
Other mineral mining	0.3	-1.33	0.32	0.81	0.05	0.15	1.40	1.41
Construction	9.6	0.96	0.16	0.43	0.04	2.01	-0.30	3.31
Lumber and wood	1.2	0.31	0.34	0.29	0.06	0.28	-0.82	0.44
Furniture & fixtures	0.9	3.82	0.07	-0.10	0.07	0.60	0.57	5.03
Stone, clay, and glass	1.2	1.59	0.15	0.01	-0.04	0.44	0.93	3.08
Primary metals	2.2	1.58	0.06	0.00	0.00	-0.03	0.68	2.30
Fabricated metals	3.2	1.92	0.09	0.07	0.01	0.43	0.21	2.74
Machinery	5.4	5.45	0.07	0.10	0.04	0.03	4.43	10.12
Electrical machinery	4.7	8.12	0.10	0.09	0.02	0.33	7.92	16.58
Motor vehicles	4.7	3.84	0.57	0.07	0.05	0.19	-0.26	4.47
Other transportation equip.	2.1	4.33	0.49	0.08	0.12	0.19	0.40	5.61
Instruments	2.3	3.02	0.01	-0.05	-0.02	0.12	-0.57	2.51
Miscellaneous mfg.	0.8	1.08	0.54	0.75	0.08	-0.01	0.10	2.54
Food	6.1	1.91	0.09	-0.02	-0.09	0.04	-0.74	1.20
Tobacco	0.6	4.15	0.54	0.60	0.03	-0.18	-8.26	-3.13
Textiles	0.9	-0.21	0.16	0.05	-0.05	-0.89	0.20	-0.75
Apparel	1.1	1.57	0.42	0.09	0.04	-1.81	-0.04	0.28
Paper	2.0	0.97	1.33	0.16	0.03	-0.16	-1.70	0.63
Printing	2.9	1.62	0.75	0.24	0.05	0.02	-1.15	1.53
Chemicals	4.7	0.33	0.17	0.44	0.05	0.04	1.07	2.11
Petroleum refining	2.4	0.72	0.38	0.04	0.03	-0.13	-0.33	0.71
Rubber and plastics	2.3	2.56	0.65	0.03	0.05	0.21	0.62	4.11
Leather	0.1	1.53	0.83	0.27	0.01	-2.12	-2.35	-1.83
Railroad transportation	0.6	0.48	0.34	0.23	-0.20	-0.80	0.74	0.79
Local & interurban transit	0.4	-1.07	0.13	0.67	0.10	1.79	1.24	2.86
Trucking & warehousing	3.0	1.53	0.12	0.28	0.05	0.92	-0.04	2.85
Water transportation	0.5	2.84	0.08	-0.08	0.02	0.32	-0.20	2.98
Air transportation	1.9	1.25	0.71	0.67	0.03	0.74	1.06	4.47
Pipelines ex. natural gas	0.1	-3.30	1.22	0.27	-0.09	-0.43	2.93	0.59
Transport services	0.6	1.59	1.47	0.19	0.04	1.14	-1.24	3.19
Telephone & telegraph	4.9	5.99	1.16	0.25	0.17	1.12	1.46	10.15
Radio and television	1.2	3.77	2.21	0.45	0.38	0.84	-4.91	2.74
Utilities	5.2	-0.06	0.23	0.02	0.26	-0.22	0.26	0.48
Wholesale trade	12.7	0.82	0.53	0.21	0.26	0.92	3.04	5.77
Retail trade	17.6	0.97	1.22	0.31	0.16	1.09	1.08	4.84
Depository institutions	5.9	0.28	1.95	-0.21	0.54	0.09	0.18	2.83
Nondepository institutions	2.0	2.40	3.92	0.79	0.03	1.51	-1.63	7.02

Table 4c
Growth Accounting
(1996-2000)

		Contributions to Output Growth						
			Capital Services					
	Domar weight	Mat.	ICT	Oth. Eqp.	Str.	Lab.	TFP	Output
Security/commodity brokers	3.7	9.49	0.34	0.20	0.08	2.70	8.76	21.57
Insurance carriers	4.3	-0.61	1.20	0.17	0.28	0.51	-1.83	-0.28
Insurance agents & brokers	1.3	1.56	0.57	0.60	0.18	0.65	-0.01	3.56
Nonfarm housing	3.1	0.03	0.00	0.00	0.04	0.12	0.00	0.10
Other real estate	8.8	0.54	0.39	0.12	0.14	0.13	0.31	1.63
Investment offices	0.5	2.93	0.84	0.38	1.16	0.38	-1.53	4.16
Hotels	1.8	1.71	0.15	0.14	0.19	0.61	-0.29	2.51
Personal services	1.4	0.77	0.14	0.06	0.12	0.38	-0.22	1.26
Business services	9.5	4.00	1.40	0.31	0.07	3.43	-0.22	8.98
Automobile services	2.1	1.02	0.11	0.59	0.04	0.76	0.45	2.96
Misc. repair services	0.8	1.84	0.38	0.18	0.04	0.11	-1.36	1.19
Motion pictures	0.6	1.35	0.29	0.06	0.14	0.93	0.68	3.45
Amusement & recreation	1.4	0.80	0.11	0.19	0.18	1.02	-0.19	2.12
Health services	7.9	1.83	0.34	0.05	0.18	0.87	-0.57	2.70
Legal services	2.2	0.56	0.32	0.02	0.05	0.52	-0.55	0.91
Educational services	0.8	1.93	0.11	-0.01	0.13	0.23	-1.24	1.15
Social services	1.0	4.70	0.26	0.01	0.03	0.74	-1.34	4.41
Membership organizations	0.8	1.38	0.13	0.01	0.02	0.69	-2.04	0.19
Other services	5.5	3.02	0.56	0.03	0.03	1.89	0.07	5.61

Table 5
Output Contribution from High-tech Capital Services and TFP, Selected Industries

	High-Tech Services			TFP		
	78-89	90-95	96-00	78-89	90-95	96-00
Nonfarm *	0.28	0.24	0.76	0.33	0.39	1.23
Machinery	0.02	0.04	0.07	2.82	1.74	4.43
Electrical machinery	0.09	0.12	0.10	2.66	5.65	7.92
Wholesale trade	0.26	0.16	0.53	1.27	1.00	3.04
Retail trade	0.22	0.42	1.22	0.01	0.03	1.08
Depository institutions	1.18	0.71	1.95	0.15	0.00	0.18
Business services	1.28	0.34	1.40	-0.34	0.85	-0.22

* Defined as private businesses excluding imputations and agriculture.

Table 6
Total Factor Productivity of Private Business by Legal Form of Organization

	1978-1989		1990-1995		1996-2000	
Total factor productivity	Corp.	NCrp.	Corp.	NCrp.	Corp.	NCrp.
Nonfarm private business	0.55	0.12	0.77	0.01	1.46	1.00
Mining	0.09	-0.68	1.12	3.32	-2.65	1.72
Manufacturing	0.95	2.25	1.00	2.32	1.27	2.77
Transportation and public utilities	0.07	0.54	0.78	2.07	0.30	-0.67
Wholesale trade	1.36	0.97	1.08	1.31	3.26	2.83
Retail trade	0.01	-0.56	0.03	-0.18	1.28	1.78
Finance, insurance, and real estate	-0.38	0.69	0.05	0.06	1.51	1.82
Services	-0.65	-0.58	-0.36	-1.05	-0.49	-1.23
Hotels and other lodging places	-1.99	1.08	1.15	-0.14	-0.46	-3.01
Personal services	-0.73	-0.30	-0.56	-0.79	-0.45	0.58
Business services	-0.49	0.58	1.18	0.91	-0.28	-4.62
Auto repair, services, and parking	-0.45	0.16	-1.31	-1.13	0.68	1.18
Miscellaneous repair services	0.26	-0.07	-1.75	-0.76	-2.05	0.01
Motion pictures	0.17	2.40	-2.80	-1.04	1.14	-2.25
Amusement and recreation	1.04	1.57	-0.29	0.85	-0.36	0.51
Health	-1.40	-1.91	-1.40	-2.16	-0.74	-1.12
Legal	-2.23	-2.10	-0.21	-1.24	-1.59	0.27
Social services	-1.09	-8.73	-0.69	-0.16	-1.53	4.29
Other services	1.02	0.15	-0.05	-2.93	0.10	1.11

Table 7
Effect of Y2K on TFP Growth

	Upper bound effect, total = $150 billion							$50 bil.
	1995	1996	1997	1998	1999	2000	Cum.	Cum.
Agriculture	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Mining and construction	0.00	0.01	0.01	0.01	0.01	0.00	0.04	0.01
Electr. equip. & instruments	0.00	0.01	0.02	0.01	0.00	-0.01	0.04	0.01
Motor vehicles and equip.	0.02	0.07	0.16	0.13	0.07	-0.06	0.39	0.13
Other durable manuf.	0.00	0.00	0.01	0.01	0.00	0.00	0.02	0.01
Chemicals & petroleum	0.00	0.01	0.02	0.02	0.01	-0.01	0.07	0.02
Other nondurable manuf.	0.00	0.01	0.01	0.01	0.00	0.00	0.03	0.01
Transportation & utilities	0.00	0.01	0.02	0.02	0.01	-0.01	0.06	0.02
Communications	0.07	0.22	0.32	0.25	0.09	-0.12	0.83	0.28
Wholesale & retail trade	0.02	0.05	0.08	0.07	0.04	-0.04	0.22	0.07
Depository & nondep. inst.	0.06	0.21	0.29	0.23	0.09	-0.13	0.75	0.25
Other finance & insurance	0.01	0.02	0.03	0.03	0.01	-0.01	0.08	0.03
Real estate	0.01	0.03	0.05	0.04	0.02	-0.02	0.13	0.04
Business & misc. prof. srvc.	0.04	0.13	0.19	0.13	0.03	-0.09	0.44	0.15
Recreation & motion pict.	0.01	0.04	0.05	0.04	0.00	-0.03	0.12	0.04
Other services	0.00	0.01	0.02	0.02	0.01	-0.01	0.06	0.02

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.[1]

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

[1] We did not adjust manufacturing in 1992 for custom software as 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.

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 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

IO code	GPO Indus.	Indc.	GPO indc.	Nmb.	SIC	Comment
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.
	65re	GP	65re	.122	6552	Ratio of empl. in 655 to 65 times ½ to split 655 b/n 6552 and 6551.
11.0103	15-17	GO	15-17	1.0	15-17	
	65re	GO	65re	.122	6552	Ratio of empl. in 655 to 65 in 1982 times ½ to split 655 b/n 6552 and 6551.
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	
	52-59	GO	542-9	.434	5462	Ratio of empl. in 5462 (bakeries) to empl. in all food stores excluding grocery stores in 1982.
18.0400	23	Sh	231-8	1.0	231-8	
	39	Dir		.1	3999pt	Shipments of 39996 (Furs dressed and dyed) in 1982 Census.
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.
	47	GP	47	.05	474, 4789pt	Assume 4741, 4738, 4785, & 4789 are same size and 4789 split evenly between 65.0100 and 65.0300.
65.0300	42	GP	42	1.0	42	Have to use value added because no gross output data available for 47.
	47		47	.025	4789pt	Same as with 65.0100.
69.0200	52-59	Clc	Mixed		52-7,9 ex. 5462	Calculated as GP(52-9)*(1-GO(548)) / GO(52-9)
	73	GP	73	.01	7396	Assumed to be small
	80	Dir		14.5	8042	Revenue of 8042 from 1982 Census.
70.0200	61	GP	61	1.0	61	
	67	GP	67	.888	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	GO	07	.140	074	Ratio of empl. in 074 to 07 in 1982.
	80	Dir			8049, 807-9	Revenue of 8049 and 807-9 from 1982 Census.

Table A1, continued
Splitting IO industries to different GPO industries, 1987, continued

IO code	GPO indus.	Indc.	GPO indc.	Nmb.	SIC	Comment
77.0504	67	GP	67	.1125	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, 89	GP	84, 89	.073	84, 8922	¼ of empl. in 873 + empl. in 84 in 1999 (from Occupation data) divided by empl. in 84, 87, and 89.
	86		86	.083	865, 9	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, 0279pt	No information so split 04.0001 evenly between 01-2 and 07-09.
	07-09			.5	071-2, 075-6, 085, 092	
11.0000	15-17	GO	15T7	1.0	15-17	
	65re	GO	653	.149	6552	Ratio of empl. in 655 to 653 in 1987 times ½ to split 655 b/n 6552 and 6551.
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	
	52-59	GO	542-9	.485	5461	Ratio of empl. in 5461 (bakeries) to empl. in all food stores excl. grocery stores in 1987.
38.0400	28	Sh	2819	1.0	2819	
	33		3334	1.0	3334	
65.0100	40	G0	40	1.0	40	
	47	G0	474-8	.375	4741, 4789pt	Assume 4741, 4738, 4785, & 4789 are same size and 4789 split evenly between 65.0100 and 65.0300.
65.0300	42	GO	42	1.0	42	
	47		474-8	.125	4789pt	Like 65.0100
69.0200	52-59	GO	527-9	1.0	52-7,9	
	52-59	GO	542-9	-.485	Ex. 5462	Have to exclude 5462, as calculated above (14.1801)
	73	Dir		.3	7396	Revenue of 7396 from 1987 Census
	80	Dir		3.5	8042	Revenue of 8042 from 1982 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.
77.0302	07-09	GO	074	1.0	074	
	80	Dir		3.6	8043, 8049	Revenue of 8043 and 8049 in 1987 Census
	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

IO code	GPO indus.	Indc.	GPO indc.	Nmb.	SIC	Comment
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	GO	110101	1.0	15, 17	
	63re	Dir		4.6	6552	Half of revenue of 6552 in 1992 Census.
11.0108	15-17	GO	110108	1.0	15, 17	
	65re	Dir		4.6	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	