

**STATISTICS OF U.S. BUSINESSES –
MICRODATA AND TABLES**

SBA/CENSUS

Data on Establishments by Firm Size

Catherine Armington, Consultant to SBA
Contract # 7.3112.0188
June 4, 1998

- 1.0 Introduction and Summary
 - 1.1 Statistics of U.S. Businesses (SUSB) defined
 - 1.2 Current availability of SUSB data
 - 1.3 Business units and their relationship
 - 1.4 Overview of SUSB microdata preparation
 - 1.5 Standard SUSB tables -- static versus dynamic data

- 2.0 Census' business register – the Standard Statistical Establishment List (SSEL)
 - 2.1 SSEL sources
 - 2.2 Company Organization Survey (COS)
 - 2.3 Census File Numbers (CFN)
 - 2.4 Employment and payroll
 - 2.5 Industry classification
 - 2.6 Identification of new and closed establishments
 - 2.7 Economic Censuses

- 3.0 County Business Patterns (CBP) – annual establishment data
 - 3.1 Selection of data from SSEL
 - 3.2 Editing of microdata and aggregate data
 - 3.3 Schedule for CBP processing

- 4.0 SUSB Tabulation file – establishment data with firm data appended
 - 4.1 Establishment data from CBP
 - 4.2 Supplemental establishment data
 - 4.3 Calculation of enterprise data
 - 4.4 SUSB static descriptive tables
 - 4.5 How SUSB establishment data differ from County Business Patterns
 - 4.6 How SUSB enterprise data differ from Census 'Enterprise Statistics'

- 5.0 SUSB longitudinal data – multi-year establishment data with firm data
 - 5.1 Construction of Longitudinal Pointer file for establishments
 - 5.2 Composite files with two years of data for each establishment
 - 5.3 SUSB dynamic change tables
 - 5.4 How dynamic change table populations differ from static tables
 - 5.5 Why dynamic changes differ from comparative statics
 - 5.6 Longitudinal Establishment and Enterprise Microdata (LEEM) at
Census' Center for Economic Studies

References

Appendix Tables and Figures

Figure 1: Data Flow into SUSB Tabulation Files

Figure 2: Data Flow into SUSB Composite Files

Tables:

- 7.1 United States: Employer Firms, Employment, Annual Payroll and Estimated Receipts by Firm Size, 1988-1995
- 7.2 Firms, Employment, Annual Payroll, and Estimated Receipts by Firm Size, SBA Region, and State, 1995
- 7.3 Firms, Establishments, Employment, Annual Payroll, and Estimated Receipts by Detailed Firm Size, 1995
- 7.4 Establishment and Employment Changes by Type of Change, 1994-1995
- 7.5 United States: Establishment and Employment Changes from Births and Deaths, 1994-1995

1.0 Introduction and Summary

1.1 Statistics of U.S. Businesses (SUSB) defined

This document described the process that the Bureau of the Census uses to construct the Statistics of U.S. Businesses (SUSB). SUSB is an annual database containing information on firms, establishments, employment, payroll, and estimated receipts, along with their industries and location (in terms of States and Metropolitan Statistical Areas). This database covers all U.S. business establishments with employees, except for railroads and households.

Since 1991 the Office of Advocacy of the U.S. Small Business Administration (SBA) has contracted with the Bureau of the Census (U.S. Dept. of Commerce) to produce comprehensive and timely data on the performance of U.S. businesses by firm size. In order to measure this, the Census Bureau builds on its annual County Business Patterns database to construct each annual Statistics of U.S. Businesses (SUSB) Tabulation file covering all business establishments with employees, and including data (firm-wide employment, payroll, estimated receipts, primary industry and State) on the firm that owns each establishment. These SUSB Tabulation files have been prepared for every year from 1988 through 1995. They are the only annual federal business data supplying information classified by firm size.

Most of the establishments in the SUSB Tabulation files have the same identification number in each annual file, as long as they remain in business. For these businesses, changes in their employment can be measured by comparing their corresponding records for different years. However, when businesses are sold, or change their legal form, or add a secondary location, their identification numbers usually change. Census has constructed a Longitudinal Pointer File to link establishment records from the SUSB Tabulation files for 1989 through 1995, so that surviving establishments can be identified even when a business changes its identification number. Using the Pointer File, business births and deaths can be accurately identified, and changes in all surviving businesses can be measured consistently.

1.2 Current availability of SUSB data

Like most other data collected or organized by the Census Bureau, these data describing individual businesses are confidential. A variety of tables have been prepared for the SBA from each of the annual files, all showing non-confidential data for establishments, aggregated by firm-size. Additional tables have been prepared from composite files, with two years of data for each establishment, to measure job generation by firms of different sizes. All of these tables are available through the Office of Economic Research at the SBA's Office of Advocacy on disks. Many of them are available on the SBA Internet site, at www.sba.gov/advo/stats. Distributions of establishments, employment, payroll and receipts by firm-size, industry, and state for 1991 through 1994 are available on CD-ROM. A new CD-ROM will include 1995, as well as 1988-1990, so that cross-sectional data for each year from 1988 through 1995 will be available for comparison.

The SUSB microdata files are constructed and maintained by the Program Research and Methods Branch of the Economic Planning and Coordination Division of the Bureau of the Census. Special tabulations of these files may be ordered from them at cost, (but the need to prevent disclosure of confidential information adds significantly to the cost). In addition, selected microdata for all establishments for 1990, 1994, and 1995 have been linked together, creating a Longitudinal Establishment and Enterprise Microdata (LEEM) file, which is available for approved research projects at Census' Center for Economic Studies. Researchers there must subscribe to "Special Sworn Status" to ensure maintenance of the confidentiality of the data, and no data (including aggregates) may be released without going through clearance procedures.

1.3 Business units and their relationship

The basic unit of SUSB data is a business establishment. An establishment is a single physical location where business is conducted or where services or industrial operations are performed. The microdata describe each establishment in terms of its employment, payroll, and estimated receipts, location (State and metropolitan area), primary industry, and legal form of organization. Additional data identify the firm (or enterprise) to which the establishment belongs, and the employment, payroll, estimated receipts, primary industry, and primary State of that firm.

Establishments that continue their operations can usually be tracked through time, even if their identification numbers are changed. Such changes result from structural, legal, or ownership changes in the business. Establishments tend to retain the same address and industry when they change ownership or legal form, and they usually retain the same name and tax identification number when they physically move their operations. Therefore it is usually possible to clearly identify the startup of a new establishment or the termination (death or closure) of an establishment, as distinguished from the appearance of a new identification number or the discontinuance of an old one.

Establishments are owned by legal entities, which are typically corporations, partnerships, or sole proprietorships. For tax purposes, each business legal entity is identified by a federal Employer Identification Number (EIN) if it has employees (or a similar Tax Identification Number (TIN) if it has no employees). Most legal entities conduct their business primarily at a single location, or establishment. Those that have multiple locations typically have only two establishments. But some legal entities own hundreds of establishments, and they may be located in different States and operate in diverse industries. When a business is sold, or changes its legal form, it becomes a new legal entity and gets a new EIN.

A firm (or enterprise or company) is the largest aggregation of business legal entities under common ownership or control. Most firms are composed of only a single legal entity which operates only a single establishment -- their establishment data and their firm data are identical. Only 4 percent of firms have more than one establishment, and they and their establishments are both described as multi-location or multi-unit. Multi-unit firms may be composed of one or more legal entities.

All SUSB establishment records include information on the firm to which the establishment belongs. The firm's nationwide employment, payroll, and receipts, along with its primary industry, and primary State are available to classify each establishment that belongs to it. This information is calculated annually for each firm by aggregating the annual data from all the establishment records associated with the firm that year.

Small multi-location firms are usually fairly simple in structure – for instance, a small store with a single branch location. Some of the larger multi-location firms are very complex and diverse, including both small and large establishments in different industries and geographic areas, organized into many different legal entities. A corporation, for instance, may own other corporations or partnerships as subsidiaries, or control others as majority shareholders or in joint ventures. A few large firms are structured with each of their establishments organized as a distinct legal entity.

Since firms are primarily legal entities, any change in their tax identification number signifies a firm death and birth. This frequently occurs while all of the establishments belonging to the firm continue their operations, so there is no economic impact from the change. Even when their legal identities remain constant, multi-location firms cannot very usefully be tracked through time because of the frequency of partial firm sales, mergers, and acquisitions. There are not yet any practical general rules for defining firm continuity in cases of complex transformations of multi-location firms. The original location (frequently the headquarters) of a multi-location firm may close while the remainder of the firm continues in business. In many cases of merger, acquisition, and divestiture, the identity of the surviving firms appears to be set primarily for public relations or tax considerations, rather than for organizational or economic continuity.

Therefore, the SUSB does not attempt to define the age or track the survival of individual firms. However, some SUSB tables do specify which new and closed establishments are/were single location firms (or original locations of multi-location firms), or are/were affiliates (or secondary locations of multi-location firms). This provides a good measure of births and deaths of firms and of secondary locations, except for the largest firm size categories, where firms often substantially survive the closing of their original location. It also misses the legal closing of many single establishment firms that are acquired by multi-location firms and continue operations as secondary locations.

1.4 Overview of SUSB microdata preparation

The Statistics of U.S. Businesses are prepared from microdata – computer-based records describing individual business locations each year. Figures 1 and 2 in the Appendix provide an overview of the flows of data from various sources which are combined to construct the SUSB data.

The Census Bureau first assembles data from a number of sources to construct an annual Standard Statistical Establishment List (SSEL). The SSEL serves Census both as a basic business name and address register for use in drawing samples and organizing business censuses, and as the basis for annual statistics on the distribution of business

establishments and their employment and payroll. Each SSEL incorporates data from the Internal Revenue Service (IRS) Business Master File (for names and addresses of business tax filers) and IRS Form 941 (for payroll and employment reported with Social Security tax payments), as well as information from Census' annual Company Organization Survey (COS) of establishments in multi-unit enterprises. Missing payroll data are imputed from prior year reported data, or other currently reported payroll data, and missing employment data are imputed from the (more frequently) reported first quarter payroll data.

Census' County Business Patterns (CBP) program selects data from the SSEL for all businesses which had any payroll payments during the year, and it further edits the key data for those, to ensure they are reported consistently with the previous year's data. CBP tables are then compiled and published showing the distributions of establishments, with their employment and payroll, by industry, State, and county.

Each annual SUSB Tabulation file is based on records drawn from the edited CBP microdata file. These include data on each establishment's state and county, industry, annual and first quarter payroll, and employment in the March 12 pay period. These establishment records are supplemented with estimates of annual receipts and codes for Metropolitan Statistical Areas. Values for the employment, payroll, receipts, industry, and primary State of the firm owning each establishment are calculated by aggregating the corresponding values from all of the establishments in each firm, and these values are appended to the record for each establishment. This SUSB Tabulation file is then tabulated to produce the annual SUSB static tables describing the business population. When a table cell covers very few businesses, the number of establishments may be provided, but any further information about those establishments will be suppressed to prevent disclosure of confidential information.

SUSB Composite files are constructed with data on each establishment for two different years, using a Longitudinal Pointer file to track the identity of each establishment in the Tabulation files for different years. The Longitudinal Pointer file is constructed by linking records for establishments in each new SUSB Tabulation file to those in the file from the previous year, matching first on Census Identification Number. If this does not match, then a variety of other characteristics are used to search for continuing establishments. This facilitates tracking individual establishments whose identification numbers have been changed due to changes in ownership or legal form or structure. Use of the Longitudinal Pointer file prevents the double counting of establishments that exist under more than one identity during a year, and it greatly reduces the incidence of false births and deaths occurring as a result of identification number changes. The SUSB dynamic change tables are produced from these Composite files which have two years of data for each establishment.

1.5 Standard SUSB tables -- static versus dynamic data

All standard tables prepared for the SBA from the Statistics of U.S. Businesses (SUSB) provide aggregated data for business establishments classified by firm-size. For these tables, firm-size in each year is measured by the aggregate nation-wide employment

of the firm to which each establishment belongs. Because employment is defined only for the pay period including March 12 each year, some of the standard tables show active firms (with positive annual payroll) reporting both firm-employment and establishment-employment of zero.

Static data provide a “snapshot” of the business population for the year. Each of the standard static tables (such as Tables 7.1 through 7.3 attached below and discussed in section 4.4) describes the business population in terms of the distribution of its establishments, firms, employment, annual payroll, and estimated receipts. Most of the static tables also provide data classified according to the industry and location of each establishment. Establishments are included in these static aggregate data if they have any positive annual payroll for the year, just as in CBP tables.

The counts of firms in these static tables indicate the number of different firms represented within the scope of each geographic and/or industry cell covered by the table, although they are classified by their firm employment size nationwide. Thus, in a state table by industry and enterprise size, the number of ‘firms’ in each cell would include all of the single-location firms in that state and industry and all of the multi-location firms that have any establishments within that state and industry. Therefore, if tables for two industries were added together, the sum of the numbers of firms would double-count every firm which operated in both industries. Only in the U.S. all-industry total tables are the counts of firms equivalent to the total number of firms in the country.

Data from these static tables for different years can be compared (comparative statics) to show how the firm-size distributions change over time. These differences measure a combination of the net changes (births less deaths) in the population of establishments, the net changes in the surviving establishments in each cell, and the net changes in classification of the surviving establishments. These classification changes may be due to changes in the industry or location of the establishment, changes in the size of the firm which still owns the establishment, or an ownership change to a different firm. Therefore, such comparisons of static tables cannot be used to measure the relative performance of businesses of different sizes, or job creation by firm size. Indeed, the more dynamic businesses are quite likely to shift to different firm-size classes over the years.

Dynamic data explicitly measure employment changes in various types of surviving establishments, as well as births, deaths, and survival in the population of establishments. The standard SUSB tables of dynamic data (Tables 7.3 and 7.4, below) are based on SUSB Composite files, which have two years of data for each establishment. These dynamic change tables focus on businesses which are active in the first quarter of the year, so they exclude establishments with no employment in the March 12 pay period. Thus businesses which start up (hire their first employee) after the first calendar quarter will not be recognized until the following year, when they have data for classification purposes. The dynamic data actually measure March to March changes, since those are the only employment data collected by the Census Bureau.

Each dynamic table includes static data describing the initial distribution of the covered businesses, combined with dynamic data on how those establishments changed (expanded, contracted, or closed) before the next year, and on new business births. For each type of change, both the gross changes in numbers of employees, and the number of establishments involved, are provided. Establishments are classified by their characteristics at the beginning of the period, except for new firm births, for which data exist only for the end of the period. When the data measuring changes are added to the data describing the distribution of businesses by firm-size at the beginning of the period, they rarely equal the static data describing businesses in the next period, because many businesses are likely to have shifted to different firm-size classes.

Firm and establishment population changes are measured in terms of the numbers of establishments which are born or which die. Births are defined as establishments with either no record or no employment in the beginning of the period, and a record with positive employment for the end of the period. Deaths are similarly defined as establishments with a record with positive employment for the beginning of the period, and either no record or no employment at the end of the period. In Table 7.5, these births and deaths are further broken out into establishments that are secondary locations (or affiliates) of multi-location firms and those that are single-location firms or original establishments of multi-location firms. These SUSB aggregate dynamic change data also summarize the components of change and calculate annual gross job generation and job destruction rates for various firm sizes and establishment industries and locations.

2.0 Census' Business Register – the Standard Statistical Establishment List (SSEL)

2.1 SSEL sources

The Standard Statistical Establishment List (SSEL) is the basic business register maintained by Census. It provides the universe from which surveys are drawn and bench-marked, as well as basic data which are periodically summarized in various publications, such as County Business Patterns. The SSEL is the underlying source of the data in the SUSB.

Each annual SSEL is actually an inter-related set of data files incorporating various types of basic data on all business establishments whose existence is recognized by the U.S. government. These are compiled from a combination of administrative data and survey responses. Data for single location firms are kept in separate files from those for establishments affiliated with multi-location firms. Data on names and addresses are organized separately from those on numerical attributes. Extensive flags are maintained to track the sources of data items and any edits that may have modified them.

The primary source of SSEL data is administrative data from the Internal Revenue Service (IRS):

- the Business Master File (BMF) provides names, addresses, and tax identification numbers for businesses that file tax returns.

- the IRS Form 941 (or 943 for farmers), for filing of Social Security Tax payments for employers, provides quarterly data on payroll and March employment for legal entities.

In addition the Census Bureau conducts an annual Company Organization Survey (COS) to collect information on establishments in multi-location firms. Non-employer data come from IRS Form 1040 Schedules C and F and Forms 1041 (estates and trusts) and 1065 (partnerships). Altogether, these cover about 246,000 multi-location firms with employees, and 9.5 million single unit businesses, including sole proprietors without employees. The multi-unit firms include roughly 1.6 million individual establishments, and each year these firms acquire or start up an additional 55,000 establishments, whose data are added to the SSEL from the COS and from other business surveys.

2.2 Company Organization Survey (COS)

The COS is conducted annually by the Bureau of the Census to collect data needed to understand the structure and identify the components of multi-location businesses. Census uses the survey responses both to identify (by name, address, company/firm number, and industry) and link together all establishments that are under common ownership, and to construct and update firm-level data. The COS' detailed data on the status, industry, payroll and employment of each location are also used to update the establishment data for multi-unit firms.

The COS is mailed to firms, and solicits information on all establishments (or locations) belonging to the surveyed firms. The response rate is 85 to 95 percent. The survey asks each company to identify establishments that have been sold, closed, started, or acquired, and those that are continuing from the prior year. For each establishment, the firm reports on first quarter and annual payroll, employment for the March 12 pay-period, and any controlling interest held by another legal entity, as well as any other business controlled by the firm.

All firms with at least 250 employees (about 30,000 firms in recent years) are surveyed annually. Medium-sized ones are surveyed on a rotating sample basis so that generally a third of them (about 50,000 in 1994, but only 20,000 in 1995) are covered in each of three years, depending on availability of funds. A new sampling scheme was introduced in 1994 to provide flexibility with minimum impact on reliability. In addition, all except tiny firms are surveyed in the Economic Census every fifth year. Tiny businesses, with less than 5 or 10 employees, depending on the industry, are not included in the COS, but are assumed to have only one establishment, unless they are identified either by another legal entity as part of their multi-unit business, or by another Census survey which incidentally identifies them as firms with multiple units. Any new small legal entities that belong to large firms should be identified promptly in the annual survey of large firms. Any new small firms created by divestitures from large firms would also be promptly identified.

There are many possible organizational structures for multi-location businesses. The most common structure for a multi-unit firm is a single legal entity (or EIN) with 2

affiliated establishments. Very large firms typically are composed of many EIN's, some of which have multiple locations and some of which are single-unit legal entities under the same ownership. There are also both large and small multi-unit firms composed of multiple EIN's with a single unit each, but this is relatively rare. Many additional legal entities function primarily as property owners or holding companies, inserted into the structure of complex businesses to own other legal entities, but having no employees themselves, so they are not covered by CBP or SUSB.

The irregularity in the sample size of the COS causes corresponding surges in the numbers of conversions from single units to multi-unit establishments. In 1990, for instance, there were about 34,900 establishment such status changes, and in 1991 this rose to 76,700. This tends to produce surges in the relative numbers of multi-unit establishments in the years with larger samples, which are primarily prior to each Economic Census and in each Economic Census (years ending with 2 or 7). Other years have correspondingly greater numbers of small and medium-sized single units which actually represent more than one location, but have reported their consolidated payroll and employment of all their locations together. This probably results in some distortion of the timing of individual establishment births, deaths, expansions, and contractions for affiliates of multi-unit firms, although the firms' overall employment changes are accurately reported each year.

2.3 Census File Numbers (CFN)

Census File Numbers (CFN's) are used to identify establishments consistently in all of Census' business files. Each CFN has 10 digits. For single unit establishments (neither owning nor owned by other establishments) the CFN is a zero followed by the nine-digit federal Employer Identification Number (EIN) of its legal entity. In these cases, the establishment, the legal entity, and the firm are identical, and have the same identification number.

Establishment records associated with multi-location firms have a completely different type of CFN, but store their EIN elsewhere in the SSEL files. Each multi-unit firm is assigned a six-digit number which Census calls an Alpha number. These Alpha numbers are randomly chosen from available numbers when their first digit is a number between 1 and 8. Those beginning with 9 have been manually assigned. For each establishment affiliated with the firm, a CFN is constructed by appending a four-digit plant or location number to the firm's Alpha number. Plant number 0001 designates the first location of a firm, and it is often an establishment which was formerly coded as a single unit, under a different CFN. It is frequently the headquarters of a multi-unit firm, except for large firms, which often build separate headquarters locations.

When a new EIN with less than 250 employees first registers with the IRS or withholds Social Security taxes for its employees, that EIN is originally assumed to represent a single location firm, and is added to the SSEL as such. If it is later found to have multiple locations itself, the entry must be revised and re-identified from being a single unit firm to being more than one unit of a new multi-unit firm. If it is found to belong to an existing multi-unit firm, the CFN of the establishment must be changed to

include that firm's Alpha number. If it actually is found to belong to another formerly single-unit firm, then both establishments must receive new CFN's including a new Alpha number representing the new multi-unit firm. Most establishment births in medium or large firms would be accurately represented in the annual SUSB files, because the SUSB data incorporate the annual COS data, and because most secondary locations share EINs with other locations in the same firm, so they can easily be properly linked to their firm.

When a multi-unit firm loses all but one of its locations, it is frequently allowed to retain its multi-unit type of identification, in the expectation that it is likely to expand to multiple locations again. Therefore there are many "multi-units" that, in fact, have only a single unit currently active in the SSEL. In the dynamic change tables, which exclude establishments with zero employment in March, there are even more multi-unit type firms which have only a single location covered, because their other locations happen not to have employees then.

2.4 Employment and payroll

Payroll includes all forms of compensation, such as salaries, wages, reported tips, commissions, bonuses, vacation allowances, sick-leave pay, employee contributions to qualified pension plans, and the value of taxable fringe benefits. It includes amounts paid to officers and executives of corporations, but does not include profit or other compensation of proprietors or partners of unincorporated businesses. The SUSB Tabulation files include both annual payroll and first quarter payroll for each establishment and firm. When first quarter payroll is zero, employment must also be zero.

Employment includes all full-time and part-time paid employees who are on the payroll in the pay period including March 12, including salaried officers and executives of corporations. Those on paid sick leave and vacation are included. Proprietors and partners of unincorporated businesses are not included. This exclusion of the management level personnel (and their profit or other compensation) from the employment counts (and payroll) of unincorporated businesses affects primarily the smallest firm-size classes, and probably reduces their apparent average compensation per employee from what it would be if all workers were included.

These March employment numbers are reported along with quarterly payroll on IRS Form 941 (or 943 for farms) for the first quarter of each year. These forms are received and posted by the IRS by mid-July each year. Although the payroll reporting is required, the employment question is voluntary, and as many as 40 to 50 percent of respondents do not provide their employment data. For these businesses, employment must be imputed from the payroll numbers reported on the same form. This imputation is based on any prior year reported employment and payroll, and on the relationship of employment to payroll for those similar businesses that do respond in the current year.

This employment imputation is relatively simple for the single establishments each representing a single-unit legal entity and firm. However, when a Form 941 represents a legal entity with multiple locations, it is much more difficult both to impute the legal entity's employment when it is not reported, and to allocate employment appropriately

among its establishments. This information is provided by the COS for all firms covered by it. In 1995 the COS provided complete employment and payroll data that was consistent with the Form 941 data for 16 percent of the multi-unit firms, covering 72 percent of their employment and 80 percent of their payroll.

For those smaller multi-unit firms that are not covered by the COS or any other survey, the employment and payroll reported for each of their EINs on Form 941 must be distributed to the various locations (multi-units) by imputation, based on any previously reported payroll or employment for individual locations, or on averages. When COS data are incomplete or inconsistent with Form 941 data, the company is often called to help work out the problems. Employment and payroll data from other Census surveys of establishments frequently provide partial information, and the remainder is imputed.

On the 1995 SSEL there were about 213,000 active multi-unit firms with 1.6 million establishments and 60 million employees. About 51,000 of these firms (with nearly a million establishments and 47 million employees) were surveyed in the 1995 COS. Due to non-response, 13 percent of these firms needed complete imputation to distribute their firm employment to their establishments, but this involved only 7 percent of the surveyed firms' employment.

The remaining 162,000 multi-unit firms had 640,000 establishments and 13 million employees. About 9 percent of their establishments were imputed and 41 percent of their employment was imputed, usually based on reported payroll and employment for the EINs in the firm and often for some of its establishments (from other surveys). Prior year employment and payroll for surviving establishments are also used. When new establishments are added to a multi-unit firm without any specific employment or payroll data, they are assigned an employment factor based on the average size of reported new establishments in their industry. For 13,000 multi-unit firms there was no basis for imputation, so zeros were assigned by default, making these firms inactive. A total of 100,000 establishments in multi-unit firms had zero payroll imputed to them, rendering them inactive also.

These various estimates of establishment employment are probably very accurate in terms of their usefulness for classifying firm-size, and for calculating aggregate employment data, since they are based on administrative data on payroll of the firms. They introduce some uncertainty into the classification of establishments by type of employment change – expanding, contracting, or stable – whenever the change is calculated from the difference between estimates, or between estimates and reports. Whether this causes an overstatement or understatement of the volatility of establishment employment is not obvious. It depends on the details of both the non-response distribution and the estimation procedures, which are very complex.

2.5 Industry Classification

The 1987 Standard Industrial Classification (SIC) system is used for classifying each establishment's primary industry to the 4-digit level. The classification is usually based on the industry description provided in its application for an Employer Identification Number

(EIN), if that was adequate. The SIC for each of the multiple units of firms is confirmed or corrected in the Company Organization Survey (COS), or in other establishment surveys. Additional SIC data (often greater detail) are supplied by matching establishment record data with Unemployment Insurance tax filing data collected from States by the Bureau of Labor Statistics (BLS). Some establishments are classified only to the 3-digit or 2-digit SIC level. When industry is not known, the SIC is coded as 9999.

Auxiliary establishments are those whose primary activity is management or support of the activities of other establishments of the same company. Their industrial classification is based on the overall activity of the company, rather than each establishment's specific support function, such as trucking, warehousing, computer processing, or management. These auxiliary establishments account for about 0.7 of the establishments and 3.3 percent of all employment.

Industry classification is verified for all surveyed establishments (in enterprises with at least 5 employees) during the quinquennial Economic Censuses. It therefore tends to change primarily in these Census years (ending in 2 or 5). However, in the year prior to the Census, information from other surveys is used most intensively to update industry wherever possible so that the correct industry survey form will be mailed out for the Census. Changes in industry classification detected during the COS, the Annual Survey of Manufactures, the Survey of Current Business, and other periodic Census surveys are used annually to update the SSEL industry classifications.

2.6 Identification of new and closed establishments

Most data on new businesses come from the IRS when tax forms are filed under new Employer Identification Numbers (EIN's) and the IRS adds them to its Business Master File of names and addresses. If these new EIN's have employment lower than the cutoff for the Company Organization Survey (COS, see above), they are assumed to represent new single-establishment enterprises, unless they belong to other multi-unit firms which include them in their response to the COS. Some of the remaining new EIN's may be included in various other business surveys and identified then as multi-units, but others may not be properly identified as multi-units until the quinquennial Economic Census.

Births of new branch locations under old EIN's for the larger multi-unit firms are picked up annually by the COS. During the CBP processing, both the COS data and other establishments with large payrolls are reviewed to identify any remaining consolidated reporting for multiple establishments, and this information is fed back to the SSEL. However, some new establishments of existing small firms may not be identified for up to 5 years. During this time their employment would appear as growth of employment in another related establishment, so the firm-size classification and the change in employment would be correct, but the employment change would be wrongly classified as expansion, rather than an affiliate birth.

An establishment is assumed to have closed if it has no payroll for two consecutive years. Single establishment records are dropped from the active SSEL if they

have no payroll for 8 quarters. For multi-unit firms the COS collects end-of-year status for each associated establishment in the survey, and those that are reported closed are flagged as deaths. However, establishments in multi-unit firms that are too small to be in the COS or to be reviewed in the CBP edits may continue to have employment allocated to them for several years as a result of imputations based on payroll of continuing establishments, and employment allocation algorithms based on prior year employment patterns within the firm. This delayed recognition of small numbers of certain types of deaths is likely to be rather closely correlated with the delayed identification of similar types of births.

2.8 Economic Censuses

Typically, the quinquennial business censuses are held in years ending with 2 and 7, and those data are released in a sequence of reports 3 to 5 years later. Forms for the 1992 censuses, for example, were mailed out to enterprises in December of 1992, requesting data covering the calendar year 1992. These forms and their follow-up forms were received back by August 1993, and were used immediately to update the 1993 SSEL data on company organization and establishment industry.

Because all surveyed businesses are asked whether they have multiple establishments, or are owned by another business, there is usually a surge of status changes in the census year, primarily small firms changing from single to multi-location status. The COS program has proposed mailing COS forms to selected 'single unit' EIN's which have shown big employment changes, in order to identify more of these status changes on an annual basis, but it has not yet been funded.

The economic census uses specialized forms specific to each industry, and collects a wide variety of detailed information on operations of the establishments and enterprises in each industry. However, enterprises with less than 5 or 10 employees, depending on industry, are not surveyed. The status and basic data for these very small businesses are derived from administrative data, primarily from the IRS.

3.0 County Business Patterns (CBP) – annual establishment data

3.1 Selection of data from SSEL

The overall picture of the data flow from the SSEL, through County Business Patterns (CBP), and onward to create each annual SUSB Tabulation file is illustrated in Figure 1 in the Appendix.

The CBP program selects data from the SSEL to produce both an extensive set of tabulations of establishment, employment, and payroll data for public use, and a carefully edited microdata file for internal Census use. Its coverage is limited to private sector non-farm establishments with employees (as evidenced by positive annual payroll), excluding railroads and most government-owned establishments. It does include government-owned establishments such as liquor stores and wholesalers, depository institutions and credit unions, and hospitals. SSEL data for the current year are matched on CFN to

(edited) CBP data for the previous year, so that changes can be reviewed. The resulting CBP establishment edit file includes current and prior year State, county, SIC, type of organization, first quarter payroll and employment, and annual payroll.

3.2 Editing of microdata and aggregate data

A preliminary establishment edit examines all large establishments (in terms of current or prior year employment, first quarter payroll, or annual payroll) and attempts to identify and correct any errors in their data, including the reporting of multiple locations consolidated into one record and the flagging of duplicate records. Current SSEL data for employment, payroll, industry, and geo-coding, and for selected ratios, are compared to those from the prior year of CBP data and to historical averages.

A COS review team searches for and resolves cases where surveyed companies have changed the degree of consolidation in their reporting, thereby either adding apparently new establishments and shifting employment from other previously consolidated establishments, or dropping establishments and including their employment in the remaining consolidated establishment(s). They also examine large businesses which have not been treated as multi-unit companies and either verify that they are, in fact, single unit businesses, or correct their employment, or find their additional establishments.

Then a cell edit reviews aggregate data classified by current year state, county, 4-digit SIC industry, and employment size to identify cell values that are inconsistent with other current year data, prior year CBP data, or other historical data. Those cells that are flagged in this edit are reviewed by analysts to resolve or verify all big changes. Any corrections to the cell aggregates are also carried through to the establishment level in the micro-data.

3.3 Schedule for CBP processing

The CBP processing begins as soon as the SSEL has incorporated the COS data. For 1995, for instance, the COS would have closed out in August 1996. Another couple of months would be needed for processing those data so that the 1995 SSEL reflects the changes in corporate affiliations and distributes employment and payroll appropriately across the separate units of each multi-unit firm. Thus the basic 1995 CBP files would be extracted from the 1995 SSEL around November 1996. The various levels of CBP review need to be sequential, but the aggregate data for final tables are produced at the same time as the cell edit tables, so that corrections can be applied to both after completion of the cell review, which would ordinarily be around June of the following year. The target release month for the 1995 CBP was August 1997.

In business census years, the schedule is slightly different to facilitate the extra processing required at all stages of the flow of business data from COS through SSEL to CBP. Thus, in 1997, the pre-census 1996 COS was closed out a month earlier, in July 1997 to allow time for preparation of appropriate 1997 census mail-out forms for each type of business, and the 1996 CBP will have its release set back a month, to September 1998.

4.0 SUSB Tabulation file – establishment data with firm data appended

4.1 Establishment data from CBP

The SUSB Tabulation file for each year is derived primarily from the County Business Patterns(CBP) on-line file, selecting all private sector establishments with non-zero annual payroll except for farms (SIC 01-02), railroads (SIC 40), Postal Service (SIC 43), private households (SIC 88), large pension, health and welfare funds (SIC 6371 with at least 100 employees), and other financial funds. Also excluded when extracting data from the CBP data base are predecessor records for multi-unit establishments and any other duplicate records. Records representing establishments in the 50 States, District of Columbia, and Puerto Rico are extracted, excluding Virgin Islands and other territories.

Data fields extracted for the preliminary SUSB Tabulation file have varied somewhat over the years, but always include the Census File Number (CFN); State, county, and place codes; legal form of organization; an edited form of the original SIC code; employment; and annual payroll.

4.2 Supplemental establishment data

To reduce the incidence of records without useful SIC codes, records with blank or ‘not available’ industry codes are matched to the subsequent year SSEL, to search for newly classified SIC codes. About 15,000 additional SIC codes are generally picked up this way. However, this has occasionally picked up some inadequately edited SICs. A few records on the 1990 SUSB Tabulation file (and possibly on others) had their SIC set to 0000, rather than 9999, to indicate unknown. They have been corrected to 9999 on the LEEM (see section 5.6 below). A few (15) records on the 1995 SUSB Tabulation file had the invalid SIC of 8053, which will be corrected to 8050.

Preliminary SUSB Tabulation records with missing or incomplete legal form of organization (LFO) codes are also passed against the subsequent year SSEL and the prior year SUSB Tabulation file in search of a valid LFO. All LFO codes are modified to provide a slightly simpler LFORM1 variable as follows:

Corporation	1
Partnership	2
Sole Proprietorship	3
Non-Profit	4
Other	5
Unknown	6

Records with unknown legal form have their business names searched for clues (such as Corp., Inc., or Ptnshp. in their name) to use as the basis for classification of their legal form.

SIC codes for auxiliary establishments are re-coded (from CBP values) to their original (SSEL) values, which indicate the industry of the enterprise for which they perform services, but only to the 2-digit level. An industry division code (IND equals 01

for agriculture through 09 for services and 10 for unknown) is determined from the first two digits of the Standard Industrial Classification (SIC).

Metropolitan Statistical Area (MSA) and CMSA codes are determined from the state, county and place codes and added to the file. They are set to nines for establishments which are not in metropolitan areas.

When CBP employment is zero and annual payroll is positive, an imputed employment level is calculated, based on the reported annual payroll, average quarterly wage factors for 2-digit SIC's, and either reported date of change or number of quarters with payroll. These imputations have not been used in SUSB Tables.

Annual sales or receipts are estimated for each establishment, as a function of its annual payroll and an appropriate receipts-to-payroll factor based on averages for its size and industry derived from receipts reported for single units in the 1992 Economic Census. Use of this procedure for years other than 1992 depends on assuming that cost inflation is the same as wage inflation, and that productivity is constant. Therefore the estimated receipts data cannot be used to examine questions about changes in productivity.

4.3 Calculation of enterprise data

Enterprise (or firm) data are constructed from the establishment records affiliated with each enterprise. Establishments with single-unit status represent single-location firms, so their firm-level data can be copied directly from their establishment-level data. For multi-location firms, the data for all affiliated establishments must be aggregated to construct enterprise-level data, which are stored in a Multi-unit enterprise file. From there, they are copied onto the records for each of their affiliated establishments.

To construct the Multi-unit enterprise file, the records for all establishments with multi-unit status (excluding establishments in Puerto Rico) are extracted from the preliminary SUSB Tabulation file. The employment, payroll and receipts of all the establishments affiliated with each enterprise are accumulated to determine that enterprise's total employment, payroll and receipts. Primary state and industry (3-digit SIC) for the multi-unit enterprise are defined as those with the largest share of annual payroll. Industry is determined by identifying first the industry division with the greatest payroll, and then the three-digit SIC with the largest aggregate payroll within that industry division. Legal form of organization of the enterprise is based on that of its affiliated establishments, with the lowest valued code taking precedence. Thus, if any of the affiliated establishments is a corporation, then the whole enterprise is assumed to be a corporation. The enterprise records on this Multi-unit enterprise file have the same format as the SUSB Tabulation file, but the Alpha number in their CFN is followed by 4 zeros, and all their establishment fields are set to blanks.

4.4 SUSB static descriptive tables

All SUSB tables prepared for the SBA provide information from the SUSB Tabulation files, with all data classified by the nation-wide employment of the firm (enterprise-size). Each of the static descriptive tables shows aggregated data from a

single annual SUSB Tabulation file, providing the distribution of businesses by firm-size, industry, and location for that single year. These tables generally cover all of the private sector businesses which had positive payroll for the year (except those few in industries not covered by CBP).

The annual payroll and estimated receipts data represent all activity during the year, or whatever part of it the businesses were active. However, the employment reported is limited to that for the pay period including March 12. Usually about ten percent of the establishments with annual payroll report no employees for March. These include seasonal businesses without employment in March, businesses starting up after March or closing before March, and businesses with temporary layoffs during March.

An average of 44,000 single-establishment firms are reorganized during the year. Between 1989 and 1995, the number of midyear reorganizations identified in the SUSB has varied between 40 thousand and 49.5 thousand. These businesses appear in both CBP and the SUSB Tabulation file in both their old and their new forms, under two different CFN's, causing double counting of the number of these firms and establishments. Most of these 'reorganizations' are actually just changes in ownership or legal form, with little economic impact. Usually such reorganizations show no first quarter payroll for the new form, so that March employment will not be double-counted. However, in about 14 percent of the cases there is March employment reported for the second form, and that is double-counted in these static tables.

The counts of "firms" in these static tables measure the number of different firms represented within the scope of each table. Thus a table covering 'All industries' in the United States will provide information on the total number of firms that are active (payroll). But the separate tables for industries will each count every firm that is active in that industry, so the numbers of firms in the industry-level tables cannot be summed up to get the national 'All industries' number. Similarly, State-level firm counts measure the number of firms with any activity in that State.

All of the tables discussed below are available on the SBA Internet site (sba.gov/advo/stats), and the annual data shown in Tables 7.1 and 7.2 (attached below), plus additional detailed by industry, are available on CD-ROM, currently for 1991 through 1995. The standardized aggregate data discussed below are limited to those which have been provided annually for several years. Some of these supercede prior tables, while other tables were produced for a year or two and were discontinued for lack of interest or funding.

Tables 7.1 and 7.2 provide distributions by employment-size of firms. They show the numbers of firms, establishments, March employment, annual payroll, and estimated receipts by industry and firm-size. Data of the type in Tables 7.1 and 7.2 are usually prepared annually for the United States for all 4-digit, 3-digit, and 2-digit Standard Industrial Classification (SIC) codes, and at the industry division level. For each State there are tables for all 2-digit SIC's and industry divisions. Firm employment is divided

into 7 classes: 0, 1-4, 5-9, 10-19, 20-99, 100-499, and 500+. Subtotals are provided for all classes with less than 20 employees and with less than 500 employees firm-wide.

Data similar to Tables 7.1 and 7.2 are also available for Metropolitan Statistical Areas by industry (2-digit SIC and division) and firm size.

A similar set of aggregated data focuses on the legal form of organization, showing the numbers of firms, establishments, March employment, annual payroll, and estimated receipts for each type of legal form, by industry division and firm employment size. These data are usually prepared for the United States and for each State. Seven firm employment sizes are specified: 0, 1-4, 5-9, 10-19, 20-99, 100-499, and 500+.

More detailed firm-size classes are provided in files of data such as shown in Table 7.3. It provides the numbers of firms, establishments, March employment, annual payroll, and estimated receipts by industry and firm-size. Table 7.3 data have recently been prepared for the United States for all 4-digit, 3-digit, 2-digit SIC, and industry division levels. For each State there are data for all 2-digit SIC's and industry divisions. Firm employment is divided into 18 classes: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-74, 75-99, 100-499, 500-999, 1000-1499, 1500-2499, and 2500+. Much of the more detailed data is suppressed to prevent disclosure.

4.5 How SUSB static data differ from County Business Patterns

Because the SUSB static data are derived from the CBP data files, the coverage is virtually identical. For 1991 and 1992, the CBP was manually edited after the micro-data were transmitted to start the SUSB preparation, reducing both the establishment and employment counts in CBP, relative to SUSB. This practice has been discontinued. In all years, the SUSB preparation filters the CBP data to eliminate a few special types of records which might occur -- those without State codes, those representing large pension and other funds, and certain extra records representing duplicates or predecessors.

Year	Establishments		SUSB	Employment	
	SUSB	CBP-SUSB		CBP-SUSB	
1988	6,016,367	2,233		87,844,303	37,329
1989	6,106,922	481		91,626,094	5,109
1990	6,175,559	4		93,469,275	6,812
1991	6,200,859	- 209		92,307,559	- 6,016
1992	6,319,300	- 1,610		92,825,797	- 24,927
1993	6,401,233	2,134		94,773,913	15,531
1994	6,509,065	211		96,721,594	11,706
1995	6,612,721	497		100,314,946	19,799

The industry classification of SUSB data differs from that of CBP primarily because the first step of SUSB processing after extracting the data from the CBP system is to search for missing industry classifications, and to fill them in with more current data from the following year's SSEL. The 1992 SUSB data, for instance, shows 71,366 unclassified establishments with 45,568 employees. CBP that year showed 86,614 with

51,167 employees. As a general rule, classifications are found for about 15 thousand of the unclassified establishments from CBP each year.

There will be occasional further differences in SIC coding because the CBP editing changes a few of the original SSEL SIC codes to group tiny industries for publication, and CBP handles coding of auxiliary establishments differently (see above). The SUSB codes are generally based on the original SIC codes from the SSEL, not the special CBP publication codes. See the appendix of an annual CBP publication for details on this.

4.6 How SUSB static data differ from ‘Enterprise Statistics’

Both Enterprise Statistics (ES) and the Statistics of U.S. Businesses (SUSB) are derived from the SSEL and use the same concept of establishment and enterprise. However, the Enterprise Statistics program has historically been limited to a somewhat narrower set of industries, and it uses different handling of auxiliary establishments, sales branches of manufacturing firms, and integrated petroleum companies. Beyond the direct effect of these differences, the classification of enterprise size and enterprise industry may differ as a result of them.

Selected data on establishments from each Economic Census are aggregated to represent entire enterprises, and these enterprise-level data are summarized in the Enterprise Statistics: Company Summary publications every five years. The most recent ES publication, covering 1987, was released in June 1991. Although the underlying data on establishments, company organization, and enterprise (or company) are the same, many of the summary statistics from the ES differ substantially. There are two basic reasons for this -- differences in scope and differences in industry classification.

The industry coverage of the Enterprise Statistics (ES) program has been increasing over time. In 1987 it excluded all establishments in Agriculture, Financial services, Public administration, and Unclassified. Its coverage of Transportation, communication and public utilities and Services was partial. Primarily as a result of this more limited coverage, the 1987 ES covered establishments with 68,140,393 employees, while the earliest SUSB data for 1988 covered 87,884,303 employees, about a third more.

Both programs construct enterprise employment by aggregating the employment of all establishments belonging to each enterprise. Naturally, when an enterprise includes establishments in the sectors excluded by ES, but included by SUSB, the calculated employment, and sometimes the enterprise size classification, comes out smaller for ES than for SUSB. Using 500 employees to define large businesses, the 1987 ES showed 56.7 percent of employment in small businesses, while the 1988 SUSB showed 54.5 percent.

Both programs determine the overall industry of the enterprise by identifying first the industry division with the most annual payroll, and then the 3-digit industry within that division with the greatest payroll. However, no standard tables from the SUSB use the enterprise industry classification – all are based on the industry of each individual

establishment. Even when SUSB enterprise industry was used, there were some substantial differences from ES classifications. Some of these differences were due to the impact of establishments included in the SUSB which were not covered in the ES. Others result from some special enterprise industry classification rules that ES uses, which are described below.

Enterprise Statistics treats auxiliary establishments – those who service only other parts of the same enterprise – as adjuncts to the enterprise defined without reference to the primary activity of the auxiliary establishments. Thus the payroll of any trucking, computer service, warehousing, sales, and research facilities which serve only other establishments in their firm will not be taken into consideration when classifying the overall industry of the firm.

In addition, ES has special rules for a couple of industries. If a manufacturing enterprise includes any petroleum refining, then any payroll of oil and gas extraction establishments gets counted with the manufacturing payroll as petroleum refining. Thus ES classifies all integrated petroleum producers in manufacturing, while SUSB would classify them in mining if that payroll were greater. ES also treats manufacturers sales branches as manufacturing, for the purpose of determining primary industry division. Therefore, some large manufacturing/wholesale enterprises which SUSB would classify as wholesale, based on their relative payroll, would be classified by ES as manufacturing.

As a result of these apparently minor differences in industry classification, the small business shares of various industries in the forthcoming 1992 Enterprise Statistics will probably differ considerably from the 1992 distribution in SUSB static tables.

5.0 SUSB Longitudinal Data – multi-year establishment data with firm data

5.1 Construction of Longitudinal Pointer file for establishments

The process of linking data from consecutive annual SUSB Tabulations files to construct accurate multi-year data for each establishment is summarized in Figure 2 in the Appendix. It begins with the linkage of two years of data, and then each additional year of data is used to extend the linkages. Linked data now extend from 1989 through 1995.

The Longitudinal Pointer file is a directory for tracking each continuing establishment. It lists up to two Census File Numbers (CFN) for each year, allowing for a maximum of one midyear reorganization during each year, as well as a possible change in identity between each year's SSEL file.

The CFN is the basic Census identification number which is assigned to each new establishment. Although it is generally maintained consistently over time, a change in ownership or legal form, or a change in status between multi-unit and single-unit, will cause a change in CFN. A complex system of computerized matching of records for establishments which might have changed CFN's is used to identify continuing establishments in the SUSB and to update the longitudinal pointer file each year.

The annual updating of the Longitudinal Pointer file uses a wide variety of information to track continuing establishments that have changed CFN's. These include matching Permanent Plant Numbers (PPN's), matching on EIN's to track changes from single-unit to multi-units businesses, and statistical matching of records for single units, based on their attributes – such as name, address, zip code, and industry.

Each establishment location has a Permanent Plant Number (PPN) which identifies its physical location and industry and this PPN should not change with ownership changes. The PPN's were revised in 1988 because problems with their processing had led to some duplication, so it is difficult to consistently track establishments prior to that revision.

An Employer Identification Number (EIN) is also associated with each establishment, identifying the legal entity to which it belongs. The EIN may be unique to an establishment or may be shared by many establishments belonging to the same legal entity. Multi-unit enterprises may be composed of one or many legal entities. An establishment's EIN will change when it has a change of ownership or legal form, but not when it changes from single-unit to multi-unit. Therefore EIN matching can be especially helpful in identifying those continuing establishments if the PPN fails. When there is more than one potentially matching establishment record with the same EIN the longitudinal match system picks the establishment with matching EIN and the same 5-digit zip code. If there are still multiple eligible match locations the system picks the one with the greatest employment within the matching zip code.

When no matches are found on any of the above bases, the remaining single-unit establishment records are further processed in search of matches within each 5-digit zip code, based on either name matching or 3-digit industry and street number matching. This identifies many of the remaining independent businesses that have changed EIN (and therefore changed CFN) and have not been already tracked with a PPN.

Using the production of the 1992-93 update to the Longitudinal Pointer file as a typical example, the linkage process began with about 6.3 million establishment records in each year's SUSB Tabulation file. A corresponding name and address file was constructed for each, containing the attributes used for statistical matching. The residual from each match step was passed to the next match step. Here are the results:

- 5,564,000 record pairs matched on CFN,
- 32,000 of the remainder matched on PPN,
- 2,500 additional matched on EIN.

The remaining unmatched multi-unit records represent 110,000 deaths in the 1992 file and 86,000 births in the 1993 file.

The remaining single-unit records were passed on to further match processes, grouped by their 5-digit zip codes. Records showing no quarterly payroll in the last quarter were considered as potential matches with other records (usually limited to those

showing no quarterly payroll in the first quarter of the same year). The results were as follows:

- matching on business name
 - 19,100 matched across years
 - 25,300 matched within 1992 (midyear reorganization)
 - 24,300 matched within 1993 (midyear reorganization)
- matching on industry (3-digit SIC) and street number
 - 10,600 matched across years
 - 11,400 matched within 1992 (midyear reorganization)
 - 12,500 matched within 1993 (midyear reorganization)

Many of the remaining records were not eligible to match because they lacked the crucial data for determining a match. These all became births and deaths by default, although with more complete information some might have been matched:

- not eligible to match because of missing or invalid zip code or industry
 - 129,000 single unit records in 1992
 - 95,000 single unit records in 1993
- not eligible to match because of non-unique zip/SIC/Street #
 - 10,500 single unit records in 1992
 - 23,100 single unit records in 1993

These and the other remaining unmatched records represent 402,000 single unit deaths in the 1992 file and 531,000 single unit births in the 1993 file.

The results of this sequence of matches were used to extend the Longitudinal Pointer file to 1993, adding the CFN's for all qualifying newly discovered 1992 mid-year reorganizations, the 1993 CFN values, and all qualifying 1993 mid-year reorganizations. Within-year matches that indicate a possible mid-year reorganization of a single unit (establishment) are subjected to an additional condition before they qualify to be added to the Longitudinal Pointer file – the unit must have existed for at least two years under the same EIN before reorganizing.

Unfortunately, each additional year of longitudinal data provides evidence in some cases that previously matched CFNs indicating mid-year reorganizations were errors, so those previously linked CFNs must be unlinked when updating the Longitudinal Pointer file. Specifically, in our example of updating the 1989-92 Longitudinal Pointer file with matches from the 1992-93 linkage process, some of the single establishments that were previously identified as having reorganized (changing to a new CFN) during 1992 will be found continuing under their original CFN in 1993. In each of these cases the updating process must eliminate the previous mid-year change, and figure out the appropriate alternative handling of the new establishment.

The volume of such revisions may be estimated roughly by noting that around 12,000 CFNs are changed during each update of the Longitudinal Pointer file. However, many of these would be counted twice, so perhaps 7-8,000 mid-year reorganizations from the last year of each version of the file are later found to be inconsistent with the following year's data.

The Longitudinal Pointer file for each set of years is fixed in format, with each establishment record containing space for two CFN's for each year covered by the Pointer File. The second value for each year is blank if no mid-year reorganization was detected, but the first value for each year is filled in for every year that the establishment exists. Thus, if an establishment continued under the same CFN throughout the period, its CFN would appear as the first value for each year, and blanks would appear in the space for the second value.

5.2 Composite Files with two years of data for each establishment

For each pair of years for which a dynamic change table is needed, a file with longitudinal composite establishment records is created. Each Composite File record includes two years' values for CFN, employment, annual payroll, estimated receipts, and mid-year reorganization flags. The record also contains single values (usually the earlier year's) for 4-digit Standard Industrial Classification, State code, CMSA code, Legal Form of Organization code, Metropolitan Statistical Area code, and enterprise employment size code, in addition to a record type code indicating whether the establishment was active in both years, just the later (birth), or just the earlier (death).

The Composite files are constructed to link data for continuing establishments as defined by the Longitudinal Pointer file for the two specified years. The concepts of birth, death, and continuity in a composite file are defined solely by the data for the two years covered by that file. If, for example, the Pointer file indicates that an establishment had SUSB Tabulation file records only for 1990, 1991, and 1993, then the 1989-90 composite file would classify it as a 1990 birth, the 1990-91 composite file would classify it as continuing, the 1991-92 file would show a death, the 1992-93 would see another birth, and the 1993-94 would show another death. However, a 1990-92 composite file would show a death, a 1991-93 composite file would show a continuing establishment, and it would not show up at all in a 1992-94 composite file.

An important factor in the design of the composite file is the specification of rules for choosing data from alternative SUSB Tabulation records, especially in case of mid-year reorganizations. For each of the two years there may be zero, one or two records available, corresponding to whether the establishment was inactive (no annual payroll), active with no reorganization, or subject to a mid-year reorganization. The basic principle is to use the earliest available data for classification variables (excluding employment, payroll, and receipts). Thus, for establishments which were active in the first of the two years, that year's data will be used for the composite record unless the value indicates 'not available.' Any classification variable (SIC, MSA, county) that is not available in that record may be taken from the next available record for that establishment.

The obvious exception to this principle of using the earliest available data is the restriction that employment, payroll, and receipts data must come from the appropriate year. And furthermore, in the case of mid-year reorganizations, if the first record has no employment, and employment is taken from the second record, then the CFN will be changed to that of the second record. Thus, for example, if a 1992-93 composite file

were being constructed and the record for the establishment's first 1992 CFN had zero in its employment field, then any second CFN (resulting from a mid-year reorganization) would be searched for positive employment. If positive employment were found for it, then the 1992 CFN on the composite file would be changed to the second CFN and that establishment's 1992 employment would be used. The same procedure would be followed for the 1993 employment and CFN.

The firm-size code is chosen according to the same general principle, but with two additional considerations. If the establishment is not active in the first period and is born as a multi-unit in the second period, then its enterprise size code is the first period size of the enterprise with the same firm identification (Alpha), if that firm existed in the first period. If there is a mid-year reorganization and the employment reported for the first CFN of the establishment was zero, and that of the second CFN was not zero, then in addition to using this second CFN and its employment, its firm-size code is also used.

In cases of mid-year reorganization, the annual payroll and annual receipts data for the year are constructed by adding together the corresponding data from the records representing the pre-reorganization form and the post-reorganization form, so that the entire year is covered.

5.3 SUSB dynamic change tables

Dynamic change tables are often called "job generation" tables. They provide measures of the business population at a point in time, associated with measures of various types of changes in those businesses which took place before the next measurement point. Thus, for instance, if comparison of the static tables for 1992 and 1993 showed a small increase in the number of establishments in services, the dynamic tables for 1992-93 would detail the sources of this increase, showing the gross deaths of 1992 establishments in each firm size, as well as the births into services by 1993. Any remaining differences must be attributed to the net transfer of establishments into or out of the industry during that year.

SUSB aggregate data such as that shown in Table 7.3 shows the numbers of establishments and their employment and their changes from births, deaths and surviving establishments, all classified by the employment size of the firm. These are available for the United States by 3-digit industry, and for each State by industry division. All size, industry, and geographic classifications are determined at the beginning of the period, except those of new establishments which are independent or born to multi-unit firms that did not exist in the beginning of the period. These latter births are classified by their ending period characteristics.

Establishment births and deaths are further broken out into those of 'enterprises' (including independent establishments and original locations of multi-unit firms) and those of 'affiliates' (additional locations affiliated with multi-unit firms) in tables such as Table 7.5. Due to the constraints of disclosure limitation, Table 7.5 type data are available only for the U.S. for the larger 6 industry divisions.

The SUSB dynamic change aggregate data cover only businesses which have employees on March 12 of the beginning or ending period. This has the disadvantage that certain seasonal businesses are never included, and businesses with temporary lack of employees in March will appear to go out of business and later start up anew. On the other hand, it has a number of important advantages. It is clean and simple – the universe is businesses with employees, and all businesses in it have employees. The employment data necessary for enterprise size calculations are always available. Births are not recognized until they have employees and information about them is then available. Deaths are recognized more promptly (the first March after they lose all employees), so their impact can be measured using their prior period data. In addition, the establishment counts are more closely associated with the employment counts (both cover first quarter only), so that calculations of average employment and payroll are more representative.

5.4 How dynamic change table populations differ from static tables

The base period number of establishments in dynamic change tables (such as Table 7.5) represents the population of businesses at a point in time, in contrast to static table establishment counts (as in Table 7.2, for instance), which include every business that existed (had a positive payroll) at any time during the year. The establishment counts in static tables include all establishments with employees on March 12, plus all those that had employees earlier that year and died by March 12, plus all those that were born later that year, plus those that are seasonal which were inactive in March. Thus, a higher business turnover rate will result in a larger difference between the static count and the dynamic count of establishments.

The static tables also double count all of the single-establishment firms which have been identified as reorganizations (changing owners or legal form, so that their Census identification numbers changed) during the year. There are around 44,000 of these each year. Most of these have no March employment in their post-reorganization form, so this double-counting has a relatively small impact on aggregate employment. However, because of their reorganization activity, most of these post-reorganization firms will be mis-classified as zero-employee businesses (along with most businesses that start up during the year). Their annual payroll will cover only the latter part of the year.

Other than these differences of definition of universe, the base period population of businesses in a dynamic table is the same as the population in a static table for the same year. For example, the Table 7.7 counts of establishments and employment for manufacturing in Alabama in 1992 will be higher than those of the Table 7.4 base period counts for 1992-93 by the number of establishments with no employment on March 12, 1992, plus any 1992 reorganizations. However, the total base period employment and payroll for manufacturing in Alabama will be nearly identical in Tables 7.2 and 7.4.

5.5 Why dynamic changes differ from comparative statics

However, it must be kept in mind that the sum of changes in the dynamic table will not usually correspond to the difference between the corresponding two static tables. Employment differences, for instance, between two years of static tables covering a sector of an economy represent a combination of (1) the net changes in employment of

the businesses that were in that sector, plus (2) the employment of new businesses starting up in that sector, plus (3) the employment of all businesses which moved into or were reclassified into the measured sector, less (4) the employment of all businesses which moved out or were reclassified out of the sector. Employment differences in the corresponding dynamic table will be limited to (1) the net changes in employment of the businesses originally operating in that sector, plus (2) the employment of the new businesses starting up in the measured sector.

To continue with the example from above, the difference between the comparable Table 7.2 employment of manufacturing establishments in Alabama for 1992 and that in 1993 will not be identical to the employment changes specified for 1992-93 in Table 7.4. The comparison of those static tables measures the net change between the years in employment of Alabama's population of manufacturing establishments. The dynamic change data in Table 7.4 measure how those manufacturing establishments that were active in Alabama in the first quarter of 1992 changed by the first quarter of 1993, in addition to providing information on the new establishments that were born into manufacturing in Alabama between March 1992 and March 1993. Some of the former establishments may have become primarily service locations, or moved into Mississippi. Other establishments may have moved in from Texas, or split into separate wholesale and manufacturing locations. Each of these actions will be reflected differently in comparisons of static tables with a dynamic table. The analyst must take care to use the measure of change which is most appropriate to his research.

5.6 Longitudinal Establishment and Enterprise Microdata (LEEM) at the Census' Center for Economic Studies

A Longitudinal Enterprise and Establishment Microdata (LEEM) file has been constructed by Census' Economic Planning and Coordination division under contract to the SBA. It is a pilot for a similar, but more extensive, longitudinal file of SUSB microdata, and was designed to support initial research on patterns of job creation and destruction. This new composite file links three years (1990, 1994, and 1995) of data for all private sector establishments with employees in any of those years. Each establishment is represented by a record which includes the start year of the establishment and three years of information extracted from the 1989-1995 Longitudinal Pointer file and from the three appropriate annual SUSB Tabulation files. For each covered year the file includes each establishment's CFNs, SIC, State, MSA, county, enterprise employment, establishment employment, and annual payroll in thousands.

More detailed documentation of the LEEM will be the first product of the research on that file, which commenced in January 1998. The pilot LEEM may be used for non-commercial research at any location of the Census' Center for Economic Studies. Each project proposal must be approved by the CES staff, and researchers must become Special Sworn Employees of the Bureau of the Census in order to maintain the confidential nature of the data. Data from other Census programs may be linked to the LEEM data (by CFN) to enrich the database for analysis.

LEEM Update

The extended Longitudinal Establishment and Enterprise Microdata (LEEM) file will continue to be a simple composite file. Each establishment record will include annual data for eight years – 1989 through 1996, unlike the preliminary LEEM file, which contained data only for 1990, 1994, and 1995. This new file was designed primarily for more detailed and longer run analyses of job generation patterns and corporate dynamics.

The primary data source for each year of economic data is the annual SBA Tabulation Files which have been produced by Census' Economic Planning and Coordination Division under contract to the SBA (also referred to as Census' Statistics of U.S. Businesses). The separate years of data for an establishment are linked together using the Longitudinal Pointer File (LPF) that was constructed to facilitate tracking individual establishments, even when they reorganize (changing identification numbers). The LEEM is a fixed format file with data for each active establishment for each year including the primary Census File Number (CFN) from the LPF, and selected variables from the SBA Tabulation File record for that CFN (plus payroll from another CFN when the LPF indicates a mid-year reorganization). An LPF start year will be appended to the establishment record, and this will be modified at the Center for Economic Studies (CES), substituting the Source Year from the 1989 Standard Statistical Establishments List (SSEL) for each LPF start year of "89". Additional specialized survey data could be linked to records in the LEEM to deepen the job generation analysis and to support research on other topics.

The handling of mid-year reorganizations – where the annual file has two records for an establishment which changed its identification sometime during the year – will be the same as it was for the preliminary LEEM.

- When there are two CFN's for a single year, we should use the data from the second (later) CFN unless that record has no employment and the first CFN has positive employment. Only if both conditions are true should the first CFN be primary.

- Classification data, such as location, industry, and enterprise employment, should be taken from the primary CFN's record, as defined above. However, if industry is missing (SIC =9999 or SIC =0000) in that record, then industry should be taken from the other CFN's record.

- The annual payroll numbers from both CFN's (part-year) records should be added together to cover the entire year.

- Only the primary CFN (defined above) should be copied from the Longitudinal Pointer file.

The basic LEEM will contain the following fields, along with their length in bytes for the ASCII file, and their type (Character or Numeric) in the compressed SAS file. For each year xx, where xx varies from 89 through 96:

10 c CFNxx	primary ID of the establishment in year xx
4 c SICxx	Standard Industrial Classification (1987) code

2 c	STATExx	Census state code
3 c	CTYxx	FIPS county code
4 c	MSAxx	Metropolitan Statistical Area code
6 n	FEMPxx	Firm (enterprise) employment
6 n	EMPxx	Establishment (CBP) employment
8 n	APAYxx	Annual establishment payroll in thousands
2 n	LPFSTARTYR	first year active in 1989-96 Longitudinal Pointer File, or source year from the 1989 SBA Tabulation File, if available.

Each record will be initialized with blanks, so that missing years of data will be blank. In SAS files a blank numeric field will be converted to a dot, while character fields will remain blank.

This would make a total of 43 characters for each of eight years, plus 2 characters for start year, or a total of 344 characters for the extended 8-year LEEM record. For the roughly 11 million establishments which might be in existence in at least one of the eight years, the full file would require about 3.8 million bytes before compression.

For transfer to the Center for Economic Studies the extended LEEM could be split into roughly 2.5 million records for establishments with data only for the period from 1989 through 1992, another 2.5 million records for establishment with data only for the period from 1993 through 1996, and the remaining 6 million records for establishments that survived across the boundary. The transferred file should be ordered by CFN89, CFN90, . . . , and CFN96.

Great interest has been expressed in use of some of the additional data attributes describing the multi-unit firms owning some of the establishments. Rather than further expanding the basic record for each year for all establishments, we would like to have separate firm files for each covered year (1989 through 1996), covering all enterprises classified as multiunit files that year. These data have already been constructed and stored in an Enterprise File for each year as part of the SBA/Statistics of U.S. Businesses data system. Data extracted from this file should include the following fields:

6 c	ALPHA	Census number identifying the firm or enterprise
5 n	NUMLOC	number of active multi-unit locations
3 c	EIC	Enterprise Industrial Code (largest 4-SIC in largest div.)
6 n	FEMP	Firm-wide employment
9 n	FAPAY	Firm-wide annual payroll
2 c	FSTATE	Primary State location (largest payroll?)

This would require a record of 31 characters for each of the multi-unit firms. There are usually less than 200,000 multi-unit firms each year, so each of these eight firm files would use less than 6.3 million bytes.

References

Office of Advocacy, U.S. Small Business Administration, and Center for Economic Studies, Bureau of the Census, Dept. of Commerce, “Longitudinal Establishment and Enterprise Microdata (LEEM) Documentation,” forthcoming in Spring 1998.

Traeger, Mitchell L. and Richard A. Moore, “Development of a Longitudinally-linked Establishment Based Register, March 1993 through April 1995,” Proceedings of the American Statistical Association Meetings, August 1995.

U.S. Dept. of Commerce, Bureau of the Census, Technical Paper 44, “The Standard Statistical Establishment List Program,” January 1979.

U.S. Dept. of Commerce, Bureau of the Census, County Business Patterns, 1992 “General Explanation” and “Appendices”.

U.S. Dept. of Commerce, Bureau of the Census, Economic Planning and Coordination Division, (internal) Transmittals of Specification/Procedures:
“Program Modifications for the SSEL Multiunit Complete Company Imputation Model,” 4/30/92.
“Sampling Specifications for the Company Organization Survey Probability Sample,” 10/13/94.
“Processing Requirements to Create a Longitudinal Data File,” 3/23/95.
“Processing Requirement for the Creation of Longitudinal Composite Records,” 7/12/96.
“Data Base Load Requirements for the Small Business Administration and International Business Machines Tabulations,” 10/10/96.

U.S. Dept. of Commerce, Bureau of the Census, Economic Surveys Division, (internal), “County Business Patterns: Specification for Tabulating the Cell Edit Universe and Identifying Cells for Analyst Review,” 4/7/88.

U.S. Dept. of Commerce, Bureau of the Census, Enterprise Statistics: Company Summary -1987. General Explanation” and “Appendices”.

Appendix Tables

Attached below

Table 7.1 United States: Employer Firms, Employment, Annual Payroll and Estimated Receipts by Firm Size, 1988-1995

	U.S., All Industries Employment Size of Firm								
	Total	0	0-4	5-9	10-19	20-99	100-499	<500	500+
Employer Firms									
1995	5,369,068	688,584	3,249,573	981,094	576,866	469,869	76,222	5,353,624	15,444
1994	5,276,964	691,141	3,208,235	964,985	563,097	452,383	73,267	5,261,967	14,997
1993	5,193,642	671,306	3,139,518	962,481	559,602	445,900	71,512	5,179,013	14,629
1992	5,095,356	644,453	3,075,280	945,802	551,912	439,084	69,156	5,081,234	14,122
1991	5,051,025	N.A.	3,036,304	941,296	551,299	439,811	68,338	5,037,048	13,977
1990	5,073,795	N.A.	3,020,935	952,030	562,610	453,732	70,465	5,059,772	14,023
1989	5,021,315	N.A.	3,003,224	937,202	553,449	443,959	69,608	5,007,442	13,873
1988	4,954,645	N.A.	2,979,905	923,580	540,988	430,640	66,708	4,941,821	12,824
Establishments									
1995	6,612,721	690,772	3,259,795	998,264	618,268	638,616	283,993	5,798,936	813,785
1994	6,509,065	693,992	3,218,076	982,695	608,804	631,324	283,782	5,724,681	784,384
1993	6,401,233	673,408	3,147,991	980,865	608,922	631,873	285,184	5,654,835	746,398
1992	6,319,300	646,065	3,082,325	964,863	606,276	634,713	283,719	5,571,896	747,404
1991	6,200,859	N.A.	3,048,830	961,391	593,302	593,248	260,595	5,457,366	743,493
1990	6,175,559	N.A.	3,032,253	970,580	599,529	590,496	254,747	5,447,605	727,954
1989	6,106,922	N.A.	3,014,009	956,347	592,901	586,494	252,335	5,402,086	704,836
1988	6,016,367	N.A.	2,989,964	943,442	583,301	581,622	244,697	5,343,026	673,341
Employment									
1995	100,314,946	0	5,395,432	6,440,349	7,734,080	18,422,228	14,660,421	52,652,510	47,662,436
1994	96,721,594	0	5,318,961	6,332,580	7,543,777	17,693,995	14,118,375	51,007,688	45,713,906
1993	94,773,913	0	5,258,195	6,313,651	7,498,345	17,420,634	13,825,238	50,316,063	44,457,850
1992	92,825,797	0	5,178,909	6,202,861	7,390,874	17,121,010	13,307,187	49,200,841	43,624,956
1991	92,307,559	0	5,151,143	6,174,730	7,386,939	17,146,411	13,143,390	49,002,613	43,304,946
1990	93,469,275	0	5,116,914	6,251,632	7,543,360	17,710,042	13,544,849	50,166,797	43,302,478
1989	91,626,094	0	5,054,429	6,152,151	7,420,196	17,353,444	13,373,640	49,353,860	42,272,234
1988	87,844,303	0	5,006,203	6,060,724	7,252,715	16,833,702	12,761,379	47,914,723	39,929,580
Annual Payroll (\$1,000)									
1995	2,665,921,824	25,787,172	141,537,925	137,083,047	175,388,093	437,065,364	361,060,815	1,252,135,244	1,413,786,580
1994	2,487,959,727	24,081,138	134,649,352	131,666,587	166,475,972	408,053,078	335,573,696	1,176,418,685	1,311,541,042
1993	2,363,208,106	22,361,727	128,968,107	127,133,193	159,153,336	385,005,072	316,183,732	1,116,443,440	1,246,764,666
1992	2,272,392,408	21,432,778	124,592,441	122,381,613	152,830,640	368,969,129	298,174,483	1,066,948,306	1,205,444,102
1991	2,145,015,851	N.A.	118,233,813	116,794,212	146,516,583	352,032,797	279,436,898	1,013,014,303	1,132,001,548
1990	2,103,971,179	N.A.	116,856,518	114,006,469	144,450,673	352,390,861	279,451,864	1,007,156,385	1,096,814,794
1989	1,989,941,554	N.A.	112,462,139	108,002,714	136,794,734	332,733,188	264,144,335	954,137,110	1,035,804,444
1988	1,858,652,147	N.A.	108,800,891	103,041,106	130,326,463	315,751,201	244,647,178	902,566,839	956,085,308
Estimated Receipts (\$1,000)									
1995	15,751,365,178	129,711,136	919,734,985	778,863,865	965,258,998	2,678,292,849	2,062,235,050	7,404,385,747	8,346,979,431
1994	14,840,452,031	122,784,184	880,763,759	752,675,397	928,380,448	2,531,402,508	1,930,758,304	7,023,980,416	7,816,471,615
1993	14,098,572,035	116,064,572	846,812,139	730,360,041	891,635,409	2,394,813,062	1,816,354,667	6,679,975,318	7,418,596,717
1992	13,605,183,510	110,778,665	820,739,417	705,146,922	859,446,404	2,292,331,108	1,717,787,820	6,395,451,671	7,209,731,839
1991	12,961,443,722	N.A.	777,737,765	680,215,533	829,024,906	2,207,714,822	1,618,186,793	6,112,879,819	6,848,563,903
1990	9,450,655,960	N.A.	626,678,614	569,639,860	681,191,539	1,715,442,516	1,316,747,423	4,909,699,952	4,540,956,008

Notes: 1990 is the first year estimated receipts were reported. Establishments are locations with active payroll in any quarter. Employment is measured in March; thus, some firms will have no employment and some annual payroll. Firms are an aggregation of all establishments owned by a parent company. This table illustrates the changing importance of firm sizes over time. It does not illustrate job growth, as firms can grow or decline and change firm size cells over time.

Source: U.S. Small Business Administration, Office of Advocacy, based on data provided by the U.S. Department of Commerce, Bureau of the Census.

Table 7.2 Firms, Employment, Annual Payroll, and Estimated Receipts by Firm Size, SBA Region, and State (Payroll and Receipts in thousands of dollars), 1995

State	Total	U.S. All Industries, Employment Size of Firm								
		0	1-4	5-9	10-19	<20	20-99	100-499	<500	500+
United States										
Firms	5,369,068	688,584	2,560,989	981,094	576,866	4,807,533	469,869	76,222	5,353,624	15,444
Establishments	6,612,721	690,772	2,569,023	998,264	618,268	4,876,327	638,616	283,993	5,798,936	813,785
Employment	100,314,946	0	5,395,432	6,440,349	7,734,080	19,569,861	18,422,228	14,660,421	52,652,510	47,662,436
Annual payroll	2,665,921,824	25,787,172	115,750,753	137,083,047	175,388,093	454,009,065	437,065,364	361,060,815	1,252,135,244	1,413,786,580
Est. receipts	15,751,365,178	129,711,136	790,023,849	778,863,865	965,258,998	2,663,857,848	2,678,292,849	2,062,235,050	7,404,385,747	8,346,979,431
Region I										
Connecticut										
Firms	78,646	8,923	37,001	14,054	8,269	68,247	6,794	1,755	76,796	1,850
Establishments	91,181	8,951	37,081	14,300	8,785	69,117	8,391	3,810	81,318	9,863
Employment	1,411,477	0	77,436	91,770	109,963	279,169	251,521	214,212	744,902	666,575
Annual payroll	47,069,731	413,442	2,149,949	2,522,069	3,126,312	8,211,772	7,634,103	6,359,809	22,205,684	24,864,047
Est. receipts	289,758,191	2,203,862	14,652,348	13,688,646	15,760,302	46,305,158	43,739,877	35,745,987	125,791,022	163,967,169
Maine										
Firms	31,689	5,102	14,571	5,355	3,043	28,071	2,335	529	30,935	754
Establishments	36,297	5,108	14,606	5,421	3,210	28,345	3,151	1,526	33,022	3,275
Employment	432,290	0	30,422	35,032	40,119	105,573	86,877	68,847	261,297	170,993
Annual payroll	9,800,256	141,051	546,633	625,416	777,171	2,090,271	1,804,573	1,384,150	5,278,994	4,521,262
Est. receipts	60,371,785	755,487	3,603,299	3,486,455	4,075,477	11,920,718	11,456,037	8,710,487	32,087,242	28,284,543
Massachusetts										
Firms	135,268	16,370	62,388	23,812	14,227	116,797	12,542	3,309	132,648	2,620
Establishments	160,341	16,405	62,576	24,143	14,988	118,112	15,511	8,055	141,678	18,663
Employment	2,735,294	0	133,676	154,876	189,221	477,773	475,760	441,935	1,395,468	1,339,826
Annual payroll	83,393,002	673,137	3,339,092	3,876,813	5,006,675	12,895,717	13,087,466	12,587,206	38,570,389	44,822,613
Est. receipts	446,215,186	3,139,357	22,463,411	21,399,715	26,493,820	73,496,303	74,957,399	63,716,103	212,169,805	234,045,381
New Hampshire										
Firms	29,989	4,025	13,370	5,436	2,967	25,798	2,608	662	29,068	921
Establishments	34,644	4,036	13,395	5,521	3,135	26,087	3,247	1,457	30,791	3,853
Employment	464,013	0	28,463	35,501	39,212	103,176	93,699	70,330	267,205	196,808
Annual payroll	11,773,897	120,786	617,371	786,623	916,264	2,441,044	2,240,858	1,693,026	6,374,928	5,398,969
Est. receipts	67,479,822	666,474	4,168,658	4,422,284	4,855,452	14,112,868	14,075,381	8,892,872	37,081,121	30,398,701
Rhode Island										
Firms	25,018	3,213	12,147	3,901	2,383	21,644	2,074	524	24,242	776
Establishments	27,765	3,213	12,163	3,952	2,483	21,811	2,392	939	25,142	2,623
Employment	379,595	0	25,256	25,504	31,636	82,396	76,309	62,133	220,838	158,757
Annual payroll	9,480,022	132,439	532,387	542,144	719,032	1,926,002	1,788,188	1,482,187	5,196,377	4,283,645
Est. receipts	52,061,207	533,390	3,546,329	3,019,934	3,782,680	10,882,333	10,400,613	7,760,088	29,043,034	23,018,173
Vermont										
Firms	18,437	2,353	8,950	3,179	1,762	16,244	1,365	316	17,925	512
Establishments	20,541	2,367	8,970	3,211	1,856	16,404	1,814	840	19,058	1,483
Employment	224,327	0	18,778	20,808	23,535	63,121	48,709	38,685	150,515	73,812
Annual payroll	5,040,562	48,804	329,349	356,984	443,585	1,178,722	1,015,655	868,142	3,062,519	1,978,043
Est. receipts	31,975,854	259,311	2,185,510	2,181,602	2,573,663	7,200,086	7,156,217	4,958,090	19,314,393	12,661,461

Notes: For state data, a firm is defined as an aggregation of all establishments owned by a parent company within a state.

Establishments are locations with active payroll in any quarter. Employment is measured in March; thus some firms will have zero employment and some annual payroll.

Source: U.S. Small Business Administration, Office of Advocacy based on data provided by the U.S. Department of Commerce, Bureau of the Census.

Table 7.3 Employer Firms, Employment, Annual Payroll and Estimated Receipts by Detailed Firm Size Classes, 1995

Data Item	Data				
State	United States	United States	United States	United States	United States
Industry	Total	Total	Total	Total	Total
SIC code	0001	0001	0001	0001	0001
FC_Desc	Firms	Establishments	Employment	Ann. Payroll	Est. Receipts
Total	5,369,068	6,612,721	100,314,946	2,665,921,824	15,751,365,178
Firm employment size classes					
0	688,584	690,772	-	25,787,172	129,711,136
1_4	2,560,989	2,569,023	5,395,432	115,750,753	790,023,849
5_9	981,094	998,264	6,440,349	137,083,047	778,863,865
10_14	381,812	403,389	4,464,009	100,253,522	542,338,443
15_19	195,054	214,879	3,270,071	75,134,571	422,920,555
0_19	4,807,533	4,876,327	19,569,861	454,009,065	2,663,857,848
20_24	119,246	137,712	2,601,165	60,389,264	348,613,127
25_29	78,971	94,650	2,119,955	49,429,530	293,211,798
30_34	57,856	72,775	1,843,120	43,284,928	262,243,590
35_39	42,308	55,837	1,560,484	36,468,831	223,852,511
40_44	33,266	45,842	1,393,752	32,792,550	207,925,449
45_49	26,488	37,919	1,242,438	29,692,486	188,918,947
50_74	75,221	121,727	4,530,582	108,709,909	679,757,940
75_99	36,513	72,154	3,130,732	76,297,866	473,769,487
100_499	76,222	283,993	14,660,421	361,060,815	2,062,235,050
0_499	5,353,624	5,798,936	52,652,510	1,252,135,244	7,404,385,747
500_999	7,566	91,039	5,228,246	136,686,110	718,582,751
1000_1499	2,592	51,258	3,167,391	86,003,796	449,198,617
1500_2499	2,110	64,638	4,052,904	113,753,822	618,478,064
2500+	3,176	606,850	35,213,895	1,077,342,852	6,560,719,999

Source: U.S. Small Business Administration, Office of Advocacy, based on data provided by the U.S. Department of Commerce, Bureau of the Census.

Table 7.4 Establishment and Employment Change by Type of Change and Firm Size, 1994-1995

Data Item	Firm Size	Data
State		U.S.
SIC description		Total, All Industries
SIC code		1
Beginning year		1994
End year		1995
Establishments		
Beginning Year	Total	5,770,090
	1_4	2,518,825
	5_9	980,828
	10_19	607,104
	20_99	627,603
	100_499	278,039
	500+	757,691
	<500	5,012,399
Net Change	Total	108,320
	1_4	74,990
	5_9	14,381
	10_19	1,069
	20_99	(3,140)
	100_499	3,485
	500+	17,535
	<500	90,785
Births	Total	695,657
	1_4	447,590
	5_9	85,375
	10_19	38,150
	20_99	30,592
	100_499	19,588
	500+	74,362
	<500	621,295
Deaths	Total	587,337
	1_4	372,600
	5_9	70,994
	10_19	37,081
	20_99	33,732
	100_499	16,103
	500+	56,827
	<500	530,510
With employment expansions	Total	1,769,311
	1_4	584,948
	5_9	323,178
	10_19	232,967
	20_99	262,375
	100_499	110,656
	500+	255,187
	<500	1,514,124
With employment contractions	Total	1,472,703
	1_4	338,518
	5_9	332,602
	10_19	229,374
	20_99	225,669
	100_499	90,766
	500+	255,774

Table 7.4 Establishment and Employment Change by Type of Change and Firm Size, 1994-1995

Data Item	Firm Size	Data
	<500	1,216,929
Employment		
Beginning Year	Total	96,687,346
	1_4	5,311,360
	5_9	6,325,466
	10_19	7,537,382
	20_99	17,685,901
	100_499	14,113,903
	500+	45,713,334
	<500	50,974,012
Net Change	Total	3,591,567
	1_4	1,028,947
	5_9	360,576
	10_19	288,260
	20_99	576,417
	100_499	359,105
	500+	978,262
	<500	2,613,305
Births	Total	5,763,600
	1_4	805,926
	5_9	547,501
	10_19	494,514
	20_99	910,089
	100_499	698,805
	500+	2,306,765
	<500	3,456,835
Deaths	Total	(4,530,705)
	1_4	(650,979)
	5_9	(447,775)
	10_19	(444,847)
	20_99	(859,025)
	100_499	(609,196)
	500+	(1,518,883)
	<500	(3,011,822)
With employment expansions	Total	10,593,050
	1_4	1,302,175
	5_9	936,448
	10_19	985,529
	20_99	2,068,486
	100_499	1,497,688
	500+	3,802,724
	<500	6,790,326
With employment contractions	Total	(8,234,378)
	1_4	(428,175)
	5_9	(675,598)
	10_19	(746,936)
	20_99	(1,543,133)
	100_499	(1,228,192)
	500+	(3,612,344)
	<500	(4,622,034)

Source: U.S. Small Business Administration, Office of Advocacy, based on data provided by the U.S. Department of Commerce, Bureau of the Census.

Table 7.5 United States: Establishment and Employment Changes from Firm Births and Deaths, 1994-1995

Data Type	Total	U.S., All Industries Employment Size of Firm (1994)						
		1-4	5-9	10-19	20-99	100-499	500+	< 500
Establishments								
Establishments in 1994	5,770,090	2,518,825	980,828	607,104	627,603	278,039	757,691	5,012,399
Changes Due to Establishment:								
Employment expansions	1,769,311	584,948	323,178	232,967	262,375	110,656	255,187	1,514,124
Employment contractions	1,472,703	338,518	332,602	229,374	225,669	90,766	255,774	1,216,929
Births, new firms (original locations)	594,369	447,091	84,782	37,023	22,741	2,482	250	594,119
Births, existing firms (secondary locations)	101,288	499	593	1,127	7,851	17,106	74,112	27,176
Deaths, original locations	497,246	370,769	68,630	33,042	21,759	2,674	372	496,874
Deaths, secondary locations	90,091	1,831	2,364	4,039	11,973	13,429	56,455	33,636
Establishments in 1995	5,878,410	2,593,815	995,209	608,173	624,463	281,524	775,226	5,103,184
Change	108,320	74,990	14,381	1,069	-3,140	3,485	17,535	90,785
Total (Net Change)	1.9	3.0	1.5	0.2	(0.5)	1.3	2.3	1.8
Due to Births	12.1	17.8	8.7	6.3	4.9	7.0	9.8	12.4
Due to Deaths	(10.2)	(14.8)	(7.2)	(6.1)	(5.4)	(5.8)	(7.5)	(10.6)
Employment in Establishments								
Employment in Establishments in 1994	96,687,346	5,311,360	6,325,466	7,537,382	17,685,901	14,113,903	45,713,334	50,974,012
Changes Due to Establishment:								
Employment expansions	10,593,050	1,302,175	936,448	985,529	2,068,486	1,497,688	3,802,724	6,790,326
Employment contractions	8,234,378	428,175	675,598	746,936	1,543,133	1,228,192	3,612,344	4,622,034
Births, new firms (original locations)	3,322,001	803,500	544,323	488,330	823,180	390,123	272,545	3,049,456
Births, existing firms (secondary locations)	2,441,599	2,426	3,178	6,184	86,909	308,682	2,034,220	407,379
Deaths, original locations	2,822,627	648,100	441,102	427,350	747,930	369,105	189,040	2,633,587
Deaths, secondary locations	1,708,078	2,879	6,673	17,497	111,095	240,091	1,329,843	378,235
Employment in Establishments in 1995	100,278,913	6,340,307	6,686,042	7,825,642	18,262,318	14,473,008	46,691,596	53,587,317
Change	3,591,567	1,028,947	360,576	288,260	576,417	359,105	978,262	2,613,305
Total (Net Change)	3.7	19.4	5.7	3.8	3.3	2.5	2.1	5.1
Due to births	6.0	15.2	8.7	6.6	5.1	5.0	5.0	6.8
Due to deaths	(4.7)	(12.3)	(7.1)	(5.9)	(4.9)	(4.3)	(3.3)	(5.9)
Due to employment expansions	11.0	24.5	14.8	13.1	11.7	10.6	8.3	13.3
Due to employment contractions	(8.5)	(8.1)	(10.7)	(9.9)	(8.7)	(8.7)	(7.9)	(9.1)

Notes: Represents activity from March 1994 to March 1995. Longitudinal data for private establishments active (payroll) in the first quarter of the year. (Establishments with no employment in the first quarter were excluded). New firm births are classified by their employment size at the first quarter. Continuing firms are establishments that began before the period and ended alive (ownership changes are considered continuing firms).

Source: U.S. Small Business Administration, Office of Advocacy, from data provided by the U.S. Department of Commerce, Bureau of the Census.