

---

## The New Business Information Tracking Series

*Alicia Robb, Office of Economic Research, Office of Advocacy, U.S. Small Business Administration  
and UNC-Chapel Hill*

---

**H**istorically, research on U.S. business activities over time has been hampered by the lack of accurate and comprehensive longitudinal data. To improve this situation, the U.S. Small Business Administration (SBA) contracted with the Bureau of the Census to develop better methods of producing firm-size data beginning in 1991. The development of a new longitudinal file with data on establishments and the firms that own them has been a joint project of the Census Bureau and SBA's Office of Advocacy since 1996. The Business Information Tracking Series (BITS) file (formerly called the Longitudinal Establishment and Enterprise Microdata (LEEM) file) currently consists of records on almost all U.S. establishments with positive payroll for 1989 through 1996.<sup>1</sup> Data for additional years will normally follow at a lag of 2 years. This data base is available to approved researchers at the Bureau of the Census's Center for Economic Studies.

This tremendously rich data source opens up numerous possibilities for research on businesses in the U.S. economy. It is the first nationwide high-quality longitudinal data base that covers the majority of employer businesses from all sectors of the economy. The BITS file contains the entire universe of private sector establishments with positive payroll, excluding farms (SIC 01-02), railroads (SIC 40), Postal Service (SIC 43), private households (SIC 88), and large pension, health, and welfare funds (SIC 6371 with at least 100 employees).<sup>2</sup> Each record contains information on an establishment for all years that it was in existence between 1989 through 1996. Some records have just 1 year of data (if the establishment existed for 1 year in that interval), while others contain data for every year from 1989 to 1996. The file is able to track an establishment over time, even through changes in ownership or legal form of organization.

The basic unit of this file is an establishment. An establishment is a physical location where a business conducts its activities. Businesses can be organized in several ways: sole proprietorship, partnership, or cor-

poration. They can have just one establishment (these are called single-unit firms) or they may have several establishments (these are referred to as multi-unit firms). Most firms are made up of just one establishment. More than two-thirds of multi-unit firms have less than four establishments, but some consist of thousands of establishments.

The annual information for each establishment includes its Census File Number (CFN),<sup>3</sup> Standard Industrial Classification (SIC),<sup>4</sup> State, metropolitan statistical area (MSA), county, place, firm employment, establishment employment, and annual payroll. Separate annual files for all multi-unit firms supplement the BITS file for 1991 through 1996. These files include firm employment, annual firm payroll, primary firm industry (at the 3-digit level), primary state of the firm (both determined by the greatest share of payroll), and the number of establishments belonging to each firm.

The primary source of the BITS file is the Standard Statistical Establishment List (SSEL) from the U.S. Bureau of the Census. This file is the Census Bureau's business register that has been maintained in some form since 1973. This section outlines the sources of the SSEL and the edits that the file goes through before reaching the next stage of the BITS processing.

Administrative records form the base of the SSEL file. The Internal Revenue Service (IRS) is one of the main sources for these records. Its Business Master File Entity (BMF) contains all business, organizational, and agricultural taxpayers on record with the IRS. Data on the location and industry of the business are used from this file. The IRS is also the source for payroll data, but this information comes from payroll tax returns. Employment as of March 12 of each year also is provided from these returns.

The Social Security Administration provides the Census Bureau information on new businesses from Form SS-4 (an application for an Employer Identifica-

tion Number (EIN) that all new businesses must fill out), which it gets from the IRS. These businesses are assigned a 4-digit standard industrial classification (SIC) code based on information on the application, as well as geographic information, estimated employment, and other indicators.

The Bureau of Labor Statistics (BLS) independently maintains its own business register, the Business Establishment List (BEL). BLS derives this list from State unemployment insurance administration records. The Census Bureau sends BLS all SSEL records that lack industry classifications for possible matches on the BEL file in order to improve Census's industry reporting. In 1996, for example, 320,000 single-unit firms were matched to BEL establishments, thus enabling Census to identify their 4-digit SIC codes. Social Security records are also used as a source of industry coding.

In addition to outside sources for the SSEL, the Census Bureau itself provides data for the file. Its *Company Organization Survey* (COS) maintains information on the organizational design and employment of multi-unit firms. This survey, conducted annually except in years ending in 2 or 7 (when economic censuses are taken), targets certain multi-unit firms which are deemed most likely to report changes in their composition, structure, or other characteristics. All multi-unit firms with more than 250 employees are surveyed every year.<sup>5</sup> However, most of those with less than 250 employees are surveyed on a rotating basis, with annual coverage depending on the availability of funds. In census years, the COS is merged with the economic census to collect more detailed information on all but some of the tiniest multi-unit firms. The COS and the economic censuses are the main source of information on multi-unit firms in the SSEL.

The economic census, done every 5 years, provides the most comprehensive updating of the register of U.S. businesses. At this time, new establishments within existing multi-unit firms are identified, and other updated information is added, such as more detailed industry identification. There are often apparent surges in the number of conversions from single-unit firms to multi-unit firms in census years due to the inability to recognize new multi-unit firms in the years between cen-

suses. In years with more limited funding, the sample of firms surveyed is often much smaller than in other years. Births of secondary establishments to multi-unit firms may be recognized later than their actual occurrence, while the job gains from these births may be incorrectly attributed to expansions of existing establishments. When the new secondary establishment is properly reported, it appears as a birth. And employment, which had been aggregated with that of another establishment, is transferred. This results in false job creation for births and matching false destruction from shrinkage. So, although the employment changes are correct for firms overall, there sometimes are distortions in the detailed makeup of the firms.

The Census Bureau produces County Business Patterns (CBP) data on an annual basis from the SSEL. These tabulations provide aggregate data on the number of establishments, employment, and payroll data for private sector nonfarm establishments with positive payroll. Since employment is measured in the pay period that includes March 12 of each year, while the payroll data represent annual payroll, it is possible for a business to have zero employment with positive payroll (for instance, if the business is seasonal or is formed after the March 12 pay period). The CBP tabulations exclude railroads and most government-owned establishments.<sup>6</sup> Each new year of data is compared to the previous year's data to check for substantial inconsistencies, and edits are done to correct for errors. In accordance with information gleaned from the COS, analysts review and correct for cases in which surveyed companies have experienced changes in their organizational structures.

The Statistics of U.S. Business (SUSB) Tabulations are annual files derived from the CBP. These files contain all private sector establishments with positive payroll excluding 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. The establishment's MSA is appended to the data record, as are updated industry classifications from the following year's SSEL and firm-level data.

Firm data are constructed for all multi-unit firms by aggregating the data from all affiliated establishments.

Single-unit firms have only one location, so their establishment data and firm data are identical. Firm employment, payroll, and receipts are calculated for multi-unit firms by summing up each over all establishments within each individual firm. Primary State and primary industry are assigned to each record using the State and industry with the largest share of annual payroll.

Most of the establishments in the SUSB Tabulation files never change identification number while they are in business. For these businesses, changes in their employment levels can be measured by comparing their corresponding records for different years. However, when a business is sold, when it changes its legal form, or when it adds a secondary location (in the case of a single-unit firm), its identification number usually changes. Census has constructed a Longitudinal Pointer file to link establishment records from the SUSB Tabulation files for 1989 through 1996, so that surviving establishments can be identified even when a business changes its identification number.<sup>7</sup> Using the Longitudinal Pointer File, establishment births and deaths can be more accurately identified. Changes in surviving establishments can be consistently measured. This pointer file was used to link annual data from 8 years of SUSB files to create the BITS file. The annual data in the BITS file are identical to the SUSB data, except for the exclusion of some single units which were doublecounted in the SUSB due to midyear reorganizations (generally less than 50,000 per year).

The final product of the processing described above is the BITS. This file is housed at the Center for Economic Studies (CES) in the U.S. Bureau of the Census. Due to the confidentiality of the microdata, researchers interested in using the BITS must submit a detailed proposal to CES, apply for sworn Census researcher status, and conduct their research at the center or one of their research data centers (RDC). However, extensive tabulations of these data are available from the SBA. These tabulated data are located on the SBA web site ([www.sba.gov/advo/stats](http://www.sba.gov/advo/stats)).

This section describes each of the variables in the BITS file in greater detail, as well as noting some specific details in defining or processing the variables.

The Census File Number (CFN<sub>xx</sub>)<sup>8</sup> uniquely identifies each establishment in the BITS file. For single-unit firms, it is a zero followed by a unique nine-digit identification number. For establishments in multi-unit firms, the CFN consists of a 6-digit number (referred to as the alpha code) that uniquely identifies the firm followed by a 4-digit number that uniquely identifies the establishment within that firm. The headquarters of a multi-unit firm are usually designated by '0001' for the last four digits of the CFN, although this is not always the case.

Annual establishment payroll (APAY<sub>xx</sub>) is made up of wages, salaries, reported tips, vacation allowances, sick-leave pay, bonuses, commissions, employee contributions to qualified pension plans, and compensation paid to corporate officers and executives. It does not include compensation to proprietors or partners of unincorporated businesses. The annual figure is either the sum of the four quarters of payroll or, in the cases of missing data, imputed values.

For single-unit establishments, the annual payroll is the sum of the four quarters of payroll. Quarterly payroll entries are obtained from IRS 941 reports. The Census Bureau imputes for missing quarters of payroll (however, less than 1 percent of payroll entries are imputed).<sup>9</sup> For multi-unit establishments, annual payroll is generally obtained from responses on the Annual Company Organization Survey (COS). Missing annual payroll data for multi-units are imputed using the affiliated administrative record data.

For 1994 and those years prior, two payroll entries from IRS Form 941—the Social Security Wages plus tips, and the Total Compensation—were used to compute quarterly payroll. In general, analysts from the Census Bureau selected the greater of the two entries. Social Security Wages were deficient due to the wage cap. However, Total Compensation did not include employee compensations to qualified pension plans. Beginning in 1995, Medicare wages were used in the quarterly payroll computations. These wages are consistent with the payroll definition specified above.<sup>10</sup>

Establishment Employment (EMP<sub>xx</sub>) includes full and part-time employees, salaried personnel, and per-

sons on sick leave or vacation in the pay period of March 12. In the case of sole proprietorships and partnerships, this figure does not include proprietors or partners of the business. This figure also excludes all contractors and volunteers, but does include temporary employees. While reporting payroll and employment is mandatory, the IRS does not put a lot of emphasis on the reporting of "Total Employees" on the 941 reports. This results in missing employment data for 15 percent to 18 percent of the establishments. A higher proportion of the larger EIN entities do not list employees, but much of these data can be provided from the COS.<sup>11</sup> Other data must be imputed from payroll changes. This is either derived from the prior year's reported employment and payroll, or from the ratio of employment to payroll reported from similar businesses. Employment data for most multi-unit establishments are collected by the Company Organization Survey. Other surveys and direct calls to companies provide additional information.

Firm employment (FEMPxx) is defined similarly but is aggregated over all establishments under a parent firm. For single-unit firms, firm employment and establishment employment are identical.

The Standard Industrial Classification (SICxx) code represents the primary industry of the establishment as classified by the 1987 Standard Industrial Classification system. This is usually derived from the industry description listed on the business's application for an Employer Identification Number (EIN). The COS, BLS subsequent matching, Social Security Administration, and other surveys often provide industry codes for those establishments that have not yet been classified, or additional detail for those that have industry detail only to the 2- or 3-digit level.<sup>12</sup> Codes are set to 9999 for unclassified establishments.

The state code (STATExx) represents the Census (not FIPS) code for the State in which the establishment is physically located. There are 50 States represented, as well as the District of Columbia. Puerto Rico, Guam, the U.S. Virgin Islands, and the Northern Mariana Islands are excluded.

The Metropolitan Statistical Area (MSAxx) code represents the MSA in which the establishment is physi-

cally located. There were 326 MSA's in the United States in 1995, but these definitions may change over time. Those establishments which are coded 9999 are either unclassified or are in non-MSA areas. In certain locations, such as New Jersey and the District of Columbia, there are no non-MSA areas.

The county code (CTYxx) represents the county in which the establishment is physically located. There are over 3,000 counties, which include parishes in Louisiana, the District of Columbia, independent cities, and boroughs/census areas in Alaska.

The place code (PLACExx) represents the place in which the establishment is physically located. The Census Bureau identifies more than 7,600 places, which are usually locations having more than 2,500 inhabitants.

The start year (STARTYR) for each establishment is originally recorded as the first year the establishment appeared on the 1989 to 1996 Longitudinal Pointer File. For cases where this is equal to 1989, the source year (SYR) variable from the 1989 SSEL is substituted for the start year from the longitudinal pointer file. The SYR variable represents the first year the establishment appeared in the SSEL. SSEL start years prior to 1977 were assigned a value of 77.

## ■ Industry Changes

When analyzing business over time, the assumption is often made that the industry of a business stays constant over time. Looking at the BITS file, it is possible to discern if this is a valid assumption. Considering only surviving establishments, the results show that single units are much more likely to have an SIC code change than are establishments of multi-unit firms.<sup>13</sup> There is a very high incidence of changes during the census year, as well as the years immediately before and after that year. The changes during 1992 are probably a combination of events: corrections to codes which were initially wrong, additional definition to primary industry codes, and actual changes in primary activity. Census puts extra effort into updates of SIC codes before the census in order to send the correct industry-specific census form to each business. The annual changes for the other years are probably more an understatement of ac-

tual changes that are occurring to establishments in the natural course of business. During the 1989-1996 interval, almost 25 percent of surviving single-unit establishments experienced a change in industry code, almost evenly distributed across the levels of SIC code changes (1 digit, 2 digit, 3 digit, and 4 digit). Over that same time period, about 14 percent of establishments from multi-unit firms experienced an SIC code change. The percentage of employment experiencing SIC code changes closely mirrored the percentage of establishments.

### ■ Firm Size Versus Establishment Size

In many instances, analysts have assumed that firm size is a rough approximation to establishment size and vice versa. Until now, there has been little data available to substantiate or refute these claims. In the case of single-unit firms, the establishment size and firm size are of course identical, but in the case of establishments from multi-unit firms, this is not the case. Since multi-unit firms are the source of almost one-quarter of all establishments and more than one-half of the employment, it is important to look at the degree to which establishment size and firm size differ.

Using 1996 as an example, it becomes clear that the assumption that these size measures can be used interchangeably as proxies for one another may introduce distortions into an analysis. For the smallest size classes 0, 1-4, and 5-9, three-fourths or more of the establishments are in the same establishment and firm-size classes. However, in the 10-19 and 20-49 size classes, this percentage drops into the sixties, and the 50-99 class is barely over 50 percent. The problem gets even worse in the next three larger size classes, with 41 percent of establishments in the 100-249 class, 32 percent of establishments in the 250-499 class, and only 31 percent of establishments in the 500-999 class in the same firm-size class. Thus, for establishments with 50 or more employees, one-half to two-thirds are in larger firm-size classes.

In fact, if a small business is defined as having less than 500 employees, more than 12 percent of establishments with less than 500 employees are in the large firm-size class. Not including the smallest establishments

(those with less than 5 employees), this problem is even more pronounced. Over 20 percent of these establishments are located in large firms. Aggregating establishments of large and small by SBA definitions and using establishment size to represent firm size include substantial segments of large firms in the small establishment size class. Allocating employment by establishment and firm sizes yields similar results.

### ■ Industry Information

Services, retail trade, construction, and finance, insurance, and real estate (FIRE) are the four industries with the largest number of establishments. However, their patterns of distribution across firm sizes vary quite dramatically by industry. Retail trade has almost three times as many establishments in the largest firm-size category as services. Establishments in the construction industry are concentrated in the smallest size classes, whereas manufacturing; transportation, communication, and public utilities (TCPU); wholesale; retail; and FIRE all had a significant number of establishments in the largest size class. Distributing employment by size and industry results in even more striking contrasts.

### ■ Job Generation

The BITS provides comprehensive data for the study of net and gross job flows in U.S. employer establishments. With these longitudinal data, it is also possible to distinguish the births and deaths of establishments from a change of ownership in surviving establishments,<sup>14</sup> and thus to separate out the contributions of each towards the net growth.

From 1995 to 1996, the net employment growth rate was 1.9 percent. However, the net growth rate varied across firm employment sizes as well as industries. For example, establishments in the firms with less than 20 employees had a growth rate of 7.3 percent, while those firms with more than 500 employees grew just 1.0 percent. The middle size class experienced a net loss in employment of 0.1 percent. Employment in establishments in the service industry grew 2.8 percent, while that in manufacturing fell 0.7 percent. However, employment in the smallest establishment size class grew the fastest in 'other productive' industries and slowest

in the distributive industries. Employment in the largest size class experienced the exact opposite pattern; it grew slowest in 'other productive' industries and fastest in the distributive industry.

## ■ Conclusion

Until now, research on U.S. business activities over time has been hindered by the lack of accurate and comprehensive longitudinal data. The new Business Information Tracking Series data are tremendously rich data that open up numerous possibilities for dynamic analysis of businesses in our economy. It is the first nationwide high-quality longitudinal data base that covers the majority of employer businesses from all sectors of the economy. Due to the confidential nature of these data, the file is located at the Center for Economic Studies in the U.S. Bureau of the Census. To access the data, researchers must submit an acceptable proposal to CES and become sworn Census researchers.

## ■ Current Research Projects Involving the BITS File

There are several current projects under way at the Center for Economic Studies that involve the BITS file. This section briefly describes these projects to illustrate the versatility and potential of this new data file.

Research on race, ethnicity, and gender as determinants of business growth and survival is possible through linking the 1992 *Survey of Minority-Owned Business Enterprises* (SMOBE) with the BITS file. The SMOBE is a survey of women and minority-owned businesses done every 5 years by the U.S. Bureau of the Census. It samples over 1 million businesses, oversampling women and minority-owned businesses, and tags administrative records with race, ethnicity, and gender variables.

A current project links the SMOBE data to the BITS and then tracks these businesses from 1992 through 1996, comparing the growth rates and survival rates of women-owned businesses with those of businesses owned by men, as well as minority-owned businesses with those that are non-minority-owned. This analysis is being extended by adding variables on the owner's education and experience, the amount of capital used

for startup, and many other owner and firm characteristics available from the 1992 *Characteristics of Business Owners Survey* (CBO) by the Bureau of the Census. The CBO survey collects additional detailed information from a subsample of the SMOBE population.

Job flows (creation, destruction, reallocation, and net change) differ by establishment age and size, by firm size, by industry, and by organizational structure. The BITS file provides detailed comprehensive data for analysis of these differences. A recent project examined these topics as well as measured the impact of births and deaths on net job growth. Various methods to measure job flows were also investigated. This analysis was extended to a more indepth investigation of the service sector. The BITS file was also used to investigate the volume and impact of U.S. merger and acquisition activity from 1990-1994. A subgroup of establishments from the BITS file was identified to be probable mergers and acquisitions. The characteristics of this group were compared to those of the rest of the establishments in the BITS file. Their job creation and destruction over the 4-year period and the 1-year period from 1994-1995 were also compared. A particular focus was 'boundary crossers,' establishments that belonged to small firms in 1990, but large firms in 1994. A companion study looked only at acquisitions over the period.

With recent advances in information technology, many have predicted that work done previously in cities would be moved to more idyllic locations, with communications primarily done over the Internet. A current project seeks to identify the overall effect of recent changes in information technology on the location of economic activity. The focus is to determine what kind of places, for example large metropolitan areas, suburbs, small towns, or rural areas, are benefiting from firms' adoption of information technology. Industry-level data on the adoption of information technology are merged with local industry growth data from the BITS to determine the differential geographic impact of information technology.

The BITS is a source of tremendously rich data that open up countless possibilities for further research. Researchers need to submit a proposal to CES to access these data ([www.census.gov](http://www.census.gov)).

## ■ Footnotes

- <sup>1</sup> Catherine Armington, Al Nucci, Paul Hanczaryk, and Ed Walker provided a great deal of helpful information. The views of this paper are not necessarily the views of the U.S. SBA or the Bureau of the Census. All errors are my responsibility.
- <sup>2</sup> Most large pension fund establishments have "employment" that represents pensioners receiving payments.
- <sup>3</sup> This is a ten-digit number that uniquely identifies each individual establishment.
- <sup>4</sup> This variable has industry detail at the four-digit level in most cases.
- <sup>5</sup> All firms that were defined as multi-unit firms as of the most recent economic census.
- <sup>6</sup> Some government organizations, for example, liquor stores and wholesalers, depository institutions and credit unions, and hospitals, are included.
- <sup>7</sup> See Richard Moore and Mitch Trager (1995), "Development of a Longitudinally-Linked Establishment Based Register: March, 1993 Through April, 1995." Presented at the Joint Statistical Meetings of the American Statistical Association in Lake Buena Vista, Florida.
- <sup>8</sup> XX represents year. For example, CFN89 is the CFN for 1989.
- <sup>9</sup> Imputed quarterly payroll entries are based on the average of the reported payroll entries. For example, if quarter two was missing, quarters one, three, and four are summed, and the total is divided by three. This value is entered for quarter two payroll.
- <sup>10</sup> I would like to thank Paul Hanczaryk of the U.S. Census Bureau for this definition.

- <sup>11</sup> Paul Hanczaryk provided further clarification.
- <sup>12</sup> There are ranking factors that define rules used for coding this variable.
- <sup>13</sup> This also excludes establishments with an industry code of 9999 (unclassified).
- <sup>14</sup> Since the BITS file measures employment on March 12 of each year, it will exclude some part-time seasonal businesses. In addition, since births are recognized when they begin to have employees, they have often been in existence with employment for some time before measured in March. The same occurs with deaths; the death of an establishment will be registered on the first March 12 in which they have no employees. Thus, the number of establishments and employees for each year represent businesses that had positive employment on March 12. This will differ from static tables, such as CBP, which includes all businesses that existed at any time during that year.

## ■ References

- Acs, Zoltan and Armington, Catherine (1998), "Longitudinal Establishment and Enterprise Microdata (BITS) Documentation," Center for Economic Studies, U.S. Bureau of the Census, CES Discussion Paper 98-9.
- Armington, Catherine (1998), "Statistics of U.S. Business—Microdata and Tables of SBA/Census Data on Establishment Size," Office of Advocacy, U.S. Small Business Administration, Washington, DC.
- Armington, Catherine and Robb, Alicia (1998), "Mergers and Acquisitions in the United States: 1990-1994," Office of Advocacy, U.S. Small Business Administration, Washington, DC.
- Brown, Shelton H. and Phillips, Bruce (1989), "Comparison Between Small Business Databases (USEEM) and Bureau of Labor Statistics (BLS)

Employment Data: 1978-1986," *Small Business Economics*, 1, (4), pp. 273-284.

U.S. Department of Commerce, Bureau of the Census (1992), *County Business Patterns*, "General Explanation" and "Appendices."

U.S. Department of Commerce, Bureau of the Census, *Enterprise Statistics: Company Summary—1987*, "General Explanation" and "Appendices."

U.S. Department of Commerce, Bureau of the Census, *Technical Paper 44*, 'The Standard Statistical Establishment List Program,' January 1979.

U.S. Department of Commerce, Bureau of the Census, Economic Planning and Coordination Division,

(internal) Transmittals of Specification/Procedures: "Program Modifications for the SSEL Multi-unit Complete Company Imputation Model," 4/30/92.

U.S. Department of Commerce, Bureau of the Census, "Processing Requirement for the Creation of Longitudinal Composite Records," 7/12/96.

U.S. Department 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.

Walker, Ed (1997), "The Census Bureau's Business Register: Basic Features and Quality Issues," presented at the Joint Statistical Meetings.