NEW DATA FOR DYNAMIC ANALYSIS: THE LONGITUDINAL ESTABLISHMENT AND ENTERPRISE MICRODATA (LEEM) FILE

by

Alicia Robb*
Office of Economic Research, Office of Advocacy, U.S. Small Business Administration
University of North Carolina at Chapel Hill
alicia.robb@sba.gov

CES 99-18 December, 1999

Abstract

Until now, research on U.S. business activities over time has been hindered by the lack of accurate and comprehensive longitudinal data. The new Longitudinal Establishment and Enterprise Microdata (LEEM) are tremendously rich data that open up numerous possibilities for dynamic analyses of businesses in the U.S. economy. It is the first nationwide high-quality longitudinal database 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. This paper describes the LEEM file, the variables contained on the file, and current uses of the data.

Key Words: Longitudinal establishment microdata, business size classification, net employment growth, job creation and destruction.

JEL classification: C80, D21, L20.

* I am grateful for the assistance provided by the Center for Economic Studies, U.S. Bureau of the Census. Recognition should go to Catherine Armington and Zoltan Acs, whose validation of the prototype LEEM file paved the way for this work. 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.

Introduction and LEEM Description

Historically, 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. This Longitudinal Establishment and Enterprise Microdata (LEEM) file currently consists of data on almost all U.S. establishments with positive payroll for 1989 through 1996. Data for additional years will normally follow at a lag of two years.

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 database that covers the majority of employer businesses from all sectors of the economy. The LEEM file contains the entire universe of private sector establishments with positive payroll, excluding farms (Standard Industrial Code (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)[1]. Each record contains information on an establishment for all years that it was in existence between 1989 through 1996. Some records have just one year of data (if the establishment existed for just one 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 corporation. 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.

[1] Most large pension funds establishments have "employment" that represents pensioners receiving

The annual information for each establishment includes its Census File Number (CFN)[2], Standard Industrial Classification (SIC)[3], state, metropolitan statistical area (MSA), county, place, firm employment, establishment employment, and annual payroll (see Table 1 for a file description). Separate annual files for all multi-unit firms supplement the LEEM 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 (see Table 2 for a file description).

The next section discusses the background of the LEEM, the processing used to create the file, as well as more detailed descriptions for each variable in the file.

Background of the LEEM

The primary source of data for the LEEM 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 LEEM 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. 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

---

payments.

[2] This is a ten-digit number that uniquely identifies each individual establishment.

[3] This variable has industry detail at the four-digit level in most cases.

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.[4] 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 five 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 censuses. 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 Social Security Administration provides the Census Bureau information on new businesses from Form SS-4 (an application for an Employer Identification Number (EIN) that all new businesses must fill out), which it gets from the IRS. These businesses are assigned a 4-digit standard industrial identification (SIC) code based on information on the application, as well as geographic information, estimated employment, and other indicators.

[4] All firms that were defined as multi-unit firms as of the most recent economic census.

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' 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 also are used as a source of industry coding.

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 non-farm 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[5]. 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 structure.

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 only have one location, so their establishment data and firm data are identical. Firm employment, payroll, and receipts are calculated

[5] Some government organizations, for example, liquor stores and wholesalers, depository institutions and credit unions, and hospitals, are included.

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.[6] Using the Longitudinal Pointer File, establishment births and deaths can be more accurately identified and 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 LEEM file. The annual data in the LEEM file is identical to the SUSB data, except for the exclusion of some single units which were doubled counted in the SUSB due to mid-year reorganizations (generally less than 50,000 per year).

The final product of the processing described above is the LEEM. 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 LEEM 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). Tables are available for the entire U.S., as well as by state, MSA, firm size, and industry. Annual and five-year gross and net employment changes are also tabulated there.

[6] 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.

Descriptions of Variables

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

Census File Number (CFNxx)[7]

The CFN uniquely identifies each establishment in the LEEM file. For single unit firms it is a zero followed by a nine-digit unique 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 is usually designated by '0001' for the last four digits of the CFN, although this is not always the case.

Annual Payroll (APAYxx)

Annual establishment payroll 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 4 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).[8] For multi-unit establishments, annual payroll generally is 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

[7] XX represents year. For example, CFN89 is the CFN for 1989.

[8] 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.

quarterly payroll. In general, analysts from the Census Bureau selected the greater of the two entries. Social Security Wages was 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.[9]

Establishment Employment (EMPxx)

Establishment employment includes full and part-time employees, salaried personnel, and persons 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.[10] 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 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.[11] 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)

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

[9] I would like to thank Paul Hanczaryk of the U.S. Census Bureau for this definition.
[10] However, if these temporary employees are supplied by a personal supply agency, these employees may be included under that personnel supply agency establishment.
[11] Again, Paul Hanczaryk provided further clarification for this variable definition.

Standard Industrial Classification (SICxx)

The SIC code represents the primary industry of the establishment as classified by the 1987 Standard Industrial Classification system. This is usually assigned from the industry description listed on the business' 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.[12] Codes are set to 9999 for unclassified establishments.

State Code (STATExx)

The state code 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.

Metropolitan Statistical Area (MSAxx)

The MSA code represents the MSA in which the establishment is physically located. There were 326 MSAs 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.

County Code (CTYxx)

The county code 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.

[12] There are ranking factors that define rules used for coding this variable.

Place Code (PLACExx)

The place code 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.

Start Year

The start year 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 77 were assigned a value of 77 and any SSEL value of 00 or 01 was not used.

It is important to distinguish the meaning behind the numbers of establishments for two distinct time periods. For the period of the LEEM file, 1989-1996, the totals for each year represent the total number of establishments with payroll which started in each of those years (Table 3). For the years before 1989, the numbers represent the number of establishments that started in that year and that survived until at least 1989 with the same CFN. This variable is used as a proxy for age.

There are obvious jumps in the numbers of new establishments in multi-unit firms in the census years of 1982, 1987, and 1992, representing delayed reporting of new multi-unit establishments. The numbers for the single units appear to be disturbed only in 1987, when the scope of the SSEL was expanded. One should use caution in analysis of this variable, especially for the multi-unit establishments.

CBP and LEEM Comparisons

For validation purposes, aggregate data from the LEEM file are now compared to published CBP tables. The number of establishments, March 12 employment, and annual payroll from the LEEM file are compared with those from the CBP tabulations for the years 1989-1996 (Table 4). The number of establishments in the LEEM file is consistently within one percent of establishments in the CBP. The number is slightly less than that of the CBP, due to the elimination of duplicate records for establishments with mid-year reorganizations in the CBP file. The employment numbers are even closer, with differences within five-tenths of one percent in every year, suggesting that many of the duplicate records had little, if any, employment. Annual payroll is within two-tenths of one percent, due to the handling of the mid-year reorganizations for the LEEM file. Whereas the CBP would have duplicate records for establishments that reorganized over the year, the LEEM recognizes the reorganization, eliminates the duplicate establishment and employment, but adds both of the partial year payrolls together in order to represent the annual payroll for the establishment.

When comparing the number of establishments in the LEEM versus CBP by major industry, the biggest difference that occurs is in the uncoded classification (Table 5). By using additional information from the following years of SSEL data, analysts are able to go back to prior years and code previously unclassified establishments. This procedure generally results in the classification of an additional 15,000 establishments each year. The published CBP data was not updated in a similar manner in earlier years. In 1990 and 1991 there are striking differences in the number of uncoded establishments in the LEEM and CBP files and a corresponding greater number of establishments in all of the major industries. However, new procedures in the SSEL processing drastically decreased the number of uncoded establishments in the following years, resulting in a smaller disparity between the number of uncoded establishments in the CBP and LEEM files. Although there are still a number of uncoded establishments from CBP that are transferred into the coded industries in the LEEM file, the elimination of duplicate establishments has the dominant effect, resulting in lower establishment counts in every industry on the LEEM file when compared to those in the CBP.

The comparison of CBP and LEEM employment by industry is similar, but they differ by a smaller magnitude (Table 6). Again, this smaller magnitude results from the fact that many of the duplicate records eliminated in the LEEM processing had little or no employment. There is a similar drop off after 1991 and 1992 in the difference in uncoded establishments, for reasons mentioned above. And in years following, employment was less in all industry categories, as well as in the uncoded class.

CFN Changes

A change in the CFN of an establishment is the result of one of three actions: 1) a change in ownership, 2) a change in the legal structure of the organization, or 3) a change from a single establishment firm to a multi-unit firm type or vice versa. CFN changes may alternatively be classified as follows:

1. A single-unit firm can become a different single-unit firm.
2. A single-unit firm can become part of a multi-unit firm.
3. An establishment in a multi-unit firm can become a single-unit firm.
4. An establishment in a multi-unit firm can become part of a different multi-unit firm.

The LEEM file allows investigation into the volume of these occurrences over the 1989-1996 period (Table 7). On an annual basis, the percentage of surviving establishments with any type of change range from 1.9 percent from 1992-1993 to 3.1 percent from 1991-1992. The average is close to 2 ½ percent, with the highest percentage centered on the year of the Economic Census in 1992. The percentage of employment with CFN changes is higher, ranging from 2.5 percent and 5.5 percent (Table 8). The most common type of change was a change from one single unit to another single unit. In the Census year there was a dramatic increase in the percentage change from a single unit firm to a multi-unit firm, when all single units were asked if they had any additional locations. Part of this increase is due to actual changes, but another part results from delayed reporting of secondary establishments since the prior Census. While this does not distort

the aggregate firm employment and annual payroll totals, it may introduce some distortions with regard to establishment size class, geographic location, and industry, and inflation of gross employment changes for when it is corrected in the Census year (see Appendix A for more detail).

Standard Industrial Classification (SIC) Changes

When analyzing business over time, the assumption is often made that the industry of a business stays constant. Looking at the LEEM file, it is possible to discern if this is a valid assumption (Table 9). Considering only surviving establishments, the results show that single units are much more likely to have a SIC code change than are establishments of multi-unit firms.[13] 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 actual 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 a SIC code change. The percentage of employment experiencing SIC code changes closely mirrored the percentage of establishments (Table 10). (See Appendix A for more detail).

[13] This also excludes establishments with an industry code of 9999 (unclassified).

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.

Looking at 1996 as an example, it is obvious that the assumption that these size measures can be used interchangeably as proxies for one another may introduce distortions into an analysis (Table 11A). 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. If the smallest establishments (those with less than 5 employees) are excluded, this problem is even more pronounced. Over 20 percent of establishments with 5-499 employees are located in large firms. Aggregating establishments of large and small by SBA definitions and using establishment size to represent firm size includes substantial segments of large firms in the small establishment size class.

Allocating employment by establishment and firm sizes yields similar results (Table 12A-H). In 1996, large firms employed almost 20 percent of the employment in establishments with 10-19 employees, over 26 percent of employment in establishments with 20-49 employees, over 38 percent of employment in establishments with 50-99 employees, more than 54 percent of employment in establishments with 100-249

employees, and more than 68 percent of employment in establishments with 250-499 employees (Table 12A). In fact, firms with 1,000 or more employees employed about 50 percent or more of the employment located in establishments with 100-999 employees. They even employed 6 percent and 13 percent of employment located in establishments with 1-4 and 5-9 employees, respectively. Use of firm size as a proxy for establishment size may be very misleading.

Establishments and Employment by Firm Size and Establishment Industry

Services, retail trade, construction, and finance, insurance, and real estate (FIRE) are the four industries with the largest number of establishments (Table 13A-H). However, 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 (Table 14A-H). Services, retail, and manufacturing far and away employ the greatest number of people. The 20-49 size class has a large share of the employment for almost every industry, with the largest size class obviously being the source of the largest share of employment for all industries, except mining and construction.

Job Generation

The next section deals primarily with the measurement of gross and net changes in the employment of establishments. The LEEM 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[14], and thus separate out the contributions of each towards the net growth.

From 1995 to 1996, the net employment growth rate was 1.9 percent (Table 15A). However, the net growth rate varied across firm employment sizes as well as industries.[15] 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 grew experienced the exact opposite pattern; it grew slowest in 'other productive' industries and fastest in the distributive industry.

Net employment growth is the difference between gross job creation resulting from births of new establishments and expansions of existing establishments and gross job destruction resulting from establishment deaths and contractions of continuing establishments. Most of the net growth in 1995-1996 was due to the net difference between expansions and contractions of establishments that were active in both years. The remainder was the difference between employment changes due to births and deaths. This proportion again varied across employment size classes and industries. Almost all of the net growth in manufacturing came from expansions minus contractions, whereas the growth was about evenly split between births and deaths and expansions and contractions in the service industry.

Almost 25 percent of establishments in existence in 1995 experienced employment expansions over the next year (Table 16A). Another 21 percent experienced

[14] Since the LEEM 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 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 include all businesses that existed at any time during that year.

[15] The collapsed industry divisions used in Tables 13 and 14 are as follows: Services includes transportation, communications, and public utilities; finance, insurance, and real estate; and services. Distribution includes wholesale and retail trade. Other productive includes agriservices, mining, and construction. Manufacturing is defined similarly to the other industry tables.

contractions and just fewer than 9 percent of them closed. About 11 percent of the number of 1995 establishments were added as births over the next year. The total establishment "turnover" is generally calculated as the number of births plus the number of deaths. In this case, the turnover rate was nearly 20 percent over the one-year period.

Over the period 1989-1996 employment grew by 11.5 percent (Table 17). This again varied by firm size with establishment in firms with less than 20 employees growing nearly 24 percent while those with more than 500 employees grew 9.4 percent. Services experienced the highest growth at 22.4 percent, while the manufacturing sector had a net employment loss of 6 percent.

Births of new establishments drove the net growth in the largest firm size category whereas continuing establishments were the dominant force in the employment growth in the smallest size class. Nearly 40 percent of the establishments existing in 1989 had closed by 1996. However, births of new establishments more than outweighed the deaths of existing establishments (Table 18). The rate of establishment deaths was only slightly higher for the smallest firm size class and the rate of births was actually slightly larger in the largest size class. Over 22 percent of continuing establishments expanded over this period, while less than 20 percent experienced contractions. This also varied by firm size, with only the continuing establishments in the smallest size class experiencing more expansions than contractions. The only exception to this was in the services industry.

Multi-Unit Firms versus Single Unit Firms

In addition to the LEEM file, there are six firm files for the years from 1991-1996 which contain information about each multi-unit firm: the primary industry of the firm, the number of establishments owned by the firm, firm employment, firm payroll, and primary state. There are about 200,000 multi-unit firms in each year. Most were in wholesale trade, retail trade, and services, but the majority of their employment was concentrated in services, manufacturing, and retail trade (Table 19A-F). Most of the firms had less than 250 employees, but the greatest percentage of employment was

located in the largest firm size class. The single unit firms looked somewhat similar. Most were located in services, retail, and construction, with employment concentrated in services, retail, and manufacturing. About half of the single unit firms were in the 1-4 size class, with very few in the largest size class.

Comparing single unit firms and multi-unit firms by firm size, single unit firms had slightly smaller average employment than did multi-unit firms across the board except for the largest size class (Table 20A-F). The average employment of this size class was almost three times as large for the multi-unit firms, driving the significant gap between the averages for multi-unit firms and single unit firms. In 1996, the average employment of multi-units was 340, compared to 9 for single units.

Breaking each group of firms into large and small, some notable differences emerge. In both cases the majority of firms are small. In 1996 for example this was the case for 92.4 percent of the multi-unit firms and 99.9 percent for single unit firms. For small multi-unit firms, the average employment was 71, whereas it was only 8 for small single unit firms (Tables 19A and 20A). For large multi-unit firms the average employment was 3,635 while the average for large single unit firms was 1,063. Even for multi-unit firms that had only one establishment at the time of measurement, the average employment was 51 for small multi-units and 1,403 for large multi-units (Table 21A). Small multi-unit firms commonly had just one or two establishments, whereas large firms were much more likely to have 4 or more.

Current Research Projects Involving the LEEM file

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

Women and Minority-Owned Businesses

Research on the 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 LEEM file. Employer businesses in the SMOBE sample are identified by their Employer Identification Number (EIN). The EIN was added to the LEEM specifically so that SMOBE data on employer businesses could be merged with the LEEM file.

The SMOBE is a survey of women and minority owned businesses done every five 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. It is the most comprehensive data available on women and minority-owned businesses. The most recent survey, the 1992 SMOBE, was released in 1997. The 1997 SMOBE is currently underway and will be released for public use in the year 2001.

After the SMOBE data were added to the LEEM, it was possible to track 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. Econometric models were used to test for the significance of gender, race, and ethnicity on business growth and business survival.

This analysis is being extended by adding variables on the owner's education and experience, the amount of capital used for start up, 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 information from a sub-sample of the SMOBE population. This survey contains much more detail on business and owner characteristics, such as the education and experience of the owner, the percentage of output the firm exports, type and source of financing used, and whether the business was franchised and/or home based. Unfortunately, the Census Bureau has canceled the 1997 CBO due to lack of funding. This has been one of the best sources of detailed information on women and minority-owned businesses.

Job Generation

Job flows (creation, destruction, reallocation, and net change) differ by establishment age and size, by firm size, by industry, and by organizational structure.

The LEEM file provides detailed comprehensive data for analysis of these differences. A recent project measured the impact of births and deaths on net job growth, as well as compared the various methods commonly used to measure job flows.

Another project analyzes the gross job generation of the service sector from 1989-1995 and compares it to that in manufacturing. Persistence patterns were compared with those expected if average annual creation and destruction were distributed across the business population independently of the prior year's changes. The relationship between average wages and gross job flows in services and manufacturing are also compared, providing some basis for discussing aspects of the relative quality of new jobs in these sectors.

Mergers and Acquisitions

The LEEM file was used to investigate the volume and impact of U.S. merger and acquisition activity from 1990-1994. A sub-group of the establishments from LEEM 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 LEEM file. Their job creation and destruction over the four-year period and the one-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. It was found that about half of all such boundary crossers belonged to the merger/acquisition group, and the other half belonged to rapidly growing firms. A companion study looked only at acquisitions over the period.

Information Technology and Business Location

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. This 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 LEEM to determine the differential geographic impact of information technology.

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 Longitudinal Establishment and Enterprise Microdata are tremendously rich data that open up numerous possibilities for dynamic analyses of businesses in the U.S. economy. It is the first nationwide high-quality longitudinal database 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. However, static and dynamic aggregate tabulations of the underlying microdata are available to everyone through SBA's Office of Advocacy.

References

Acs, Zoltan, and Catherine Armington, 1998, "Longitudinal Establishment and Enterprise Microdata (LEEM) Documentation," Center for Economic Studies, U.S. Bureau of the Census, CES Discussion Paper 98-9.

Acs, Zoltan, Catherine Armington, and Alicia Robb, 1999, "Measuring Job Flow Dynamics in the U.S. Economy," Center for Economic Studies, U.S. Bureau of the Census, CES Discussion Paper 99-1.

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, D.C.

Armington, Catherine, and Alicia Robb, 1998, "Mergers and Acquisitions in the United States: 1990-1994," Office of Advocacy, U.S. Small Business Administration, Washington, D.C.

Armington, Catherine, and Alicia Robb, 1999, "An Investigation into the Volume and Impact of Business Acquisition Activity in the United States: 1990-1994," Office of Advocacy, U.S. Small Business Administration, Washington, D.C.

Brown, Shelton H. and Bruce Phillips, 1989, "Comparison Between Small Business Databases (USEEM) and Bureau of Labor Statistics (BLS) Employment Data: 1978-1986," *Small Business Economics*, 1, (4), 273-284.

Kolko, Jed, 1999. "The High-Tech Rural Renaissance?: Information Technology, Firm Size, and Rural Employment Growth," Office of Advocacy, U.S. Small Business Administration, Washington, D.C.

Moore, Richard 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.

U.S. Department of Commerce, Bureau of the Census, Country Business Patterns, 1992, "General Explanation" and "Appendices."

--------, Enterprise Statistics: Company Summary- 1987, "General Explanation" and "Appendices."

--------, *Technical Paper 44*, 'The Standard Statistical Establishment List Program,' January 1979.

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.

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

Appendix A

Cautions for LEEM Users:

Changes in Firm Type and in Industry Classification in Continuing Establishments
Exhibited in Services in the 1989-1995 LEEM
Noted by Catherine Armington, October 1999

Users of economic microdata frequently classify such data by characteristics that they treat as permanent, which, in fact, change over time. This may result in errors in analysis, due to the inclusion or exclusion or misclassification that occurs as a result of such changes. The following discussion of change in characteristics of service establishments from the 1989 through 1995 LEEM file may be useful for quantifying the size of such changes, qualifying relevant analyses, and defining suitably flexible rules for classification. Although these examples are limited to the service sector, similar patterns would probably be found in other sectors, to a greater or lesser degree.

Changes between Single Unit and Multi-unit Status

The Census File Number (CFN) of an establishment identifies it as either a single-unit firm or as a primary or secondary location of a multi-unit firm (or enterprise. Since much of the information gathering and processing at Census is handled quite differently for multi-units than for single units, there is some stickiness about recognizing changes between these two types of establishments/firms. Most of these changes were expected to be concentrated in 1992 (and 1997), when the Economic Census provides more comprehensive data than in the intervening years. This proved to be the case for establishments in the service sector.

We examined the distribution of service establishments by types of changes in identity, including all service establishments that existed throughout each time interval (annual changes, and 1989 to 1995 changes). Of the service establishments that had positive payroll in both 1989 and 1990, 86.4% were single-units (or independent firms) in 1989, and 84.9% had the same single-unit establishment identification number in 1990. The 1.4% that changed ownership or legal form, while remaining single-unit establishment firms, represents well fewer than 2 percent of the single-units. This rate of reported changes in ownership of single-unit firms appears to be constant over time, without any surge associated with the Economic Census in 1992. A total of 6.1% of the service establishments that existed in both 1989 and 1995 had changed from single units to multi-unit status over that period.

Only 0.1 of all surviving 1989 service establishments changed from single-unit to multi-unit status by 1990, and this was typical of most annual change periods. However, in the 1992 Census it was found that 1.4% of surviving service establishments were classified as single in 1991 and were multi-units (part of a multi-location enterprise or firm) by 1992. These changes probably took place at a fairly constant rate between 1987 (the prior Economic Census) and 1992, but were not discovered and reported until that census.

Looking at the corresponding distribution of employment in surviving service establishments by types of changes in identity, we see that the typical annual change from single-unit to multi-unit status affected 0.6% of employment, but the change between 1991 and 1992 affected 5.7% of employment in surviving service establishments. If this change is allocated across the five-year interval between censuses, it suggests that an annual average of 1 to 2 percent of service employment is in establishments that change from single to multi-unit status. This change could be the result of acquisition by, or merger with, another firm, or the original firm's acquiring or starting up a secondary location.

Over the six-year interval from 1989-1995, changes from single to multi-unit status affected 1.8% of surviving service establishments, but this included 5.7% of the employment in these establishments. Since only 56% of the employment in surviving service establishments was in single units in 1989, those that changed status represent 15 percent of the employment in surviving service establishments that were originally single units.

This concentration of delayed changes in reporting from single-unit to multi-unit status has another, more complex, side effect on the measurement of gross employment changes. In many cases these establishments that were reported as single units in 1991 and multi-units in 1992 were actually already multiple establishments in 1991 with consolidated reporting. When their status is corrected, their reported employment may fall drastically, because some portion of their reported 1991 employment was actually in other establishments, to which it is correctly attributed in 1992. This causes a surge in reported startups of secondary locations of multi-units, and a fall in the reported employment of the primary locations (which had previously included the consolidated employment of all their secondary establishments). However, the total employment for the firm (the aggregate of all commonly owned establishments) is undisturbed by this reporting change.

Numerous complex tabulations designed to isolate the size of this problem generally showed it to affect about 400,000 employees in 1992 in services, or a little over 1 percent of the total. This means that for 1991-1992 the job destruction in single unit service establishments was overstated by 400,000, and the job creation by births of multi-unit establishments was overstated by 400,000, although the net change classified by size of firm was not affected. This problem appeared to be distributed over all sizes of establishment/firms, but was particularly concentrated in those with 100 to 250 employees.

The multi-unit service establishments changed ownership with greater frequency (1.5% out of 13.2% over 6 years) than the single unit ones (6.1% out of 86.8%). There also appears to be an increasing trend in their annual rates of ownership change. Of the 13.6% of surviving units that were in multi-unit firms in 1989, 0.3% changed ownership by 1990. This proportion increased to 0.5% for the annual periods in the mid 90's. In contrast to the single units, the Economic Census period, from 1991 to 1992, showed

fewer changes than normal among the multi-units. Over the 6-year period, 1.5% from the 13.2% of surviving establishments that were in multi-unit firms changed ownership, which is more than one of every nine.

When we look at employment, we see that over 2% of employment in the more recent annual periods was in multi-units that changed ownership to another multi-unit firm (out of the 47% of employment that was in multi-units). Over the six-year period, 43.8% of employment in surviving units was in multi-units in 1989, and 6.8% was in multi-units that changed ownership to other multi-unit firms by 1995. In combination with the single units that become multi-units during this period, a total of 14.1% of the employment in units that survive is in units that were probably acquired by another firm.

Very few multi-unit establishments get reclassified as single-unit ones, except between 1990 and 1991, which appears to be an artifact of some unusual processing of the data -–perhaps a one-time correction or purging of the multi-unit register. It affected 0.3% of the surviving service establishments, and 1.2% of their employment.

Changes in SIC codes

Changes in Standard Industrial Classification (SIC) codes may represent either corrections of codes later found to be in error, or the result of changes in the primary activity of an establishment. They also sometimes result from the acquiring of more detailed information about these activities, which allows for more detailed classification of previously correct, but less detailed, classifications. We have no basis for allocating the changes found among these various causes of change.

Surviving service establishments were distributed by SIC code change and firm type, for each available annual period and for the entire 1989 to 1995 period. This detailed table is available at CES, with the file documentation. This analysis includes all establishments that survive with employees in both years (at the beginning and ending of each period), which also have industry coded in services in at least one of the years, and not equal to 9999 (unclassified) in either year. A total of nearly 17% of both the single units and the multi-units had some type of change in industry coding between 1989 and 1995. And this change was roughly equally divided across the various levels and types of change – into, or out of services, and, within services, at each of the 4, 3, and 2-digit levels of classification. It is clearly not the case that such changes are primarily limited to the 4-digit level of detail. Only two-fifths of the changes were limited to the same 2-digit classification.

The timing of the recording of industry changes is strongly affected by the cycles of Census processing of business data. In 1991-92 there was a surge in all kinds of changes, associated with the 1992 economic census. But the rates of change were much higher in 1992-93, particularly within the 3 and 4-digit level, as additional industry detail was acquired for (primarily newly reported) establishments.

The distribution of employment in service establishments with changes in industry coding is quite similar to that of establishments, but with somewhat greater

change shares among single units. This suggests that the changes in single units tend to be in relatively large single units, while those in multi-units are in relatively small multi-units.

Similar tabulations were done to calculate the proportions of manufacturing establishments that change their industrial coding during this period. They show considerably higher overall rates of change than were found for services, and the industry coding for single unit manufacturing firms is more volatile than that for manufacturing establishments in multi-unit firms. It is particularly curious that over 6% of the single units in manufacturing in 1992 were reclassified from some other industry division in 1991. The similar distributions for employment in surviving manufacturing establishments reveal that the reclassified single units have average employment, while the reclassified multi-units tend to be relatively smaller multi-units.

Summary

For most analytical purposes these recorded changes in the characteristics of establishments, and the sometimes irregular patterns of change, are insignificant. However, researchers must be wary of focusing on measuring changes that could be substantially affected by these factors. Many problems can be avoided by more careful definition of categories. Irregularities can be smoothed out by use of annual averages across longer periods. At the worst, this information may be helpful in supplying boundaries for uncertainty in interpretation of results.