

ASSOCIATING ESTABLISHMENTS INTO ENTERPRISES
FOR A MICRODATA FILE OF THE U.S. BUSINESS POPULATION

Catherine Amington and Marjorie Odle, The Brookings Institution

I. Overview

The Brookings Small Business Microdata Project began work in January 1980 with the goal of defining and building a microdata base including all domestic American businesses. This will be used for analysis of the impact of public policy on the small business sector of the U.S. economy. A four-year project was foreseen which involved:

1. Defining the appropriate population and its relation to aggregate measures of business.
2. Integrating available microdata into a large representative sample.
3. Using other microdata sources to enrich the sample data.
4. Establishing the techniques to develop longitudinal data in order to identify changes in the business population, employment and structure.

The basic reporting unit for the data base we are developing is the business establishment, i.e., a single business location with one or more employees, usually with a single product or service. Employment data are available predominantly on an establishment basis. On the other hand, accounting conventions and other administrative procedures dictate that most other business data be reported on an enterprise (firm) basis. Alternative definitions of "small business" abound in public policy research and implementation. Though numerous committees have been formed and studies conducted to establish standards for differentiating small and large businesses, there is as yet no single accepted definition. For the purposes of this paper we shall define small businesses as firms with fewer than 100 employees.

Working on the generally held assumption that most small business firms comprise only a single establishment, the two reporting units (establishment and enterprise) frequently have been considered equivalent for small businesses. Given our stated definition, 278,000 of the 323,000 multi-establishment firms in our data base do qualify as small businesses. On the other hand, however, 11.5 percent of the establishments with fewer than 100 employees, representing 32 percent of the employment of these small establishments, are actually part of complex (multi-establishment) firms which have more than 100 employees. Therefore, even when considering only small businesses, it is necessary to carefully distinguish enterprise basis data from establishment basis data. This is especially important when comparing data for small businesses with data for large businesses. This paper explains how we established the correspondence between establishments and firms within the 1977 USEEM, our business data base, and discusses the implications of some of the new information derived from these associated data.

II. Source of Establishment Data

The 1977 U.S. Establishment and Enterprise Microdata (USEEM) file is based primarily on data from the Dun and Bradstreet Duns Market Identifiers File (DMI) from early 1979. An extract of economic and descriptive data for each establishment was taken from the DMI File, and the range and distribution of all these data were checked. Various types of errors and minor omissions were corrected. Three percent of the records lacked employment data which we subsequently estimated for them, based on medians calculated for SIC classes within each state from the 1977 County Business Patterns data. A complete description of all the changes in the data and the new structure of the file is available. ^{1/} The data base now includes 4.7 million business establishments

with complete reporting of employment figures and industry classification for both the establishment and the firm, and also age, organizational status and geographic data for each establishment. Other data, not directly relevant to this discussion, are also contained in USEEM. ^{2/}

The 1977 USEEM represents the population of domestic U.S. business establishments with employees around the end of 1977. The coverage is somewhat broader than that of the Census Bureau's County Business Patterns or that of the Unemployment Insurance program of the Department of Labor. The USEEM includes some farm establishments, numerous large semi-governmental businesses in transportation, education and health, and some large non-profit organizations, all of which are excluded from those two governmental data sources.

The original DMI File data included indicators of each establishment's organizational status (e.g., single, headquarters, subsidiary, branch) and pointers to higher level establishments in each firm. Our Multi-establishment Enterprise File (Tree File) began as an extract from the DMI File containing records for all the establishments with ownership ties to other establishments (i.e., complex establishments). The purpose of the Tree File was to provide the data necessary for in depth study of the organizational linkages presented in the DMI File. Such study was preliminary to investigation of the consistency in reporting of employment data in these complex organizations.

At the core of any study of employment data contained in the DMI File are two different reported employment figures. One of these records employment in the establishment, and the second is a more inclusive total firm employment figure. The second is reported for all establishments except branches. For single-establishment firms and establishments classified as the ultimate owner of a firm, this total employment figure represents enterprise employment - the consolidated employment for all establishments in that enterprise, including all subsidiaries and branches. Total employment reported for a subsidiary company represents the employment of the subsidiary and any branches it may have. Table 1 below shows the number of establishments and their employees according to their organizational status in the original DMI File and in USEEM.

Table 1
Number of Establishments and Employment
by Organizational Status
Before and After Tree Completion Process

| | <u>Establishments</u> | | | | |
|--------|-----------------------|---------|------------|---------|-----------|
| | Simple | Top | Subsidiary | Branch | Total |
| Before | 3,345,000 | 390,000 | 78,100 | 456,000 | 4,269,000 |
| After | 3,414,000 | 323,000 | 77,700 | 884,000 | 4,699,000 |

| | <u>Establishment Employment</u> (in thousands) | | | | |
|--------|---|--------|------------|--------|--------|
| | Simple | Top | Subsidiary | Branch | Total |
| Before | 28,900 | 12,400 | 4,984 | 19,800 | 66,100 |
| After | 30,100 | 11,900 | 4,975 | 38,500 | 85,500 |

NOTE: Figures are rounded to thousands (establishments) and hundred thousands (employees). Additional significant digits are included if necessary to make clear distinctions.

SOURCE: Version IIA of the interim file (USA3, Table 15) and Version I of USEEM (V4DMU, Table 8).

When we summed up the reported employment figures in the original file, the discrepancy between aggregate establishment employment and aggregate enterprise employment totalled 15 million employees (65 million in establishments vs 80 million in enterprises). This indicated either a large systematic error in employment reporting or deficiencies in establishment coverage. We checked the levels of employment reported for a large sample of firms and found no evidence of over-reporting. In order to analyze and subsequently eliminate the discrepancy, we needed to clarify the organizational status of each establishment and to group the establishments into enterprises. Then we could (a) complete the organizational structure of each enterprise, (b) determine any employment reporting discrepancy within that firm, (c) correct the discrepancy appropriately, and, finally, (d) reaggregate the establishment employment data for each enterprise.

III. Enterprise Structure Data in the DMI File

The Tree File originally included data for over 924,000 complex establishments. Nearly half of these establishments were branches of firms with a headquarters at a different location. Branches are usually secondary locations, frequently with a different activity or product, but wholly owned and consolidated with the headquarters for accounting purposes. A headquarters is the primary establishment in a firm which has branches. Though occupying the same location, different divisions of a company might be identified as separate branch establishments if they have the characteristics of separate businesses. Over 78,000 establishments were subsidiaries, which were separate legal entities with their own accounting system, but were majority-owned by another establishment (the parent). The other 390,000 complex establishments were presumed to be parents and headquarters. These represented both the owners of the branches and subsidiaries on the file, and also some headquarters establishments whose branches were not reported on the DMI File. The DMI File did not include foreign subsidiaries of domestic businesses, and foreign employment was not included in the total employment figures for parent companies. Furthermore, the DMI File did not include records for sales branches of manufacturing firms; however, the employees in these branches were included in total firm employment figures.

The headquarters/branch relationship is relatively simple, involving only two levels of organization. In the branch record, employment was reported for the branch location, and a code was provided to indicate that it was a branch, along with a pointer to the headquarters record. The headquarters record was coded as a headquarters (which implied that it had at least one branch establishment), but there was no indication of how many branches it had, nor were there any pointers to its branches. Each headquarters record reported a figure for its employment at that location (establishment employment), as well as total employment - which should be the sum of its own establishment employment and that of all its branches. If the data were grouped by firm, these employment figures could be compared to see if all branches had been reported for each headquarters. (See Section V.)

The parent/subsidiary relationship is more complex for two reasons. First, both subsidiary establishments and parent establishments may also be headquarters and have branches under them. Second, parents may also be subsidiaries of other parents. Occidental Petroleum, having nine levels of parent/subsidiary relationships, is an extreme example of this organizational complexity. For each subsidiary record, the DMI File reported the employment at that establishment and, if it was also a headquarters, the total employment of itself and any branches. One code in each record in the DMI File indicated whether the establishment was a headquarters, and another code showed whether it was a subsidiary. Each

subsidiary had a pointer to its parent, but there was no code to indicate whether it was itself a parent.

In order to deal more efficiently with these multilevel, complex affiliations between establishments, D & B has used the concept of the ultimate owner, the top of each enterprise structure of related establishments. Every establishment that is part of a multi-establishment enterprise in the DMI File, including the top, should have a pointer to the ultimate owner. Using these pointers, we sorted the establishment records into the enterprises to which they belonged and examined their enterprise level employment data.

The Tree File originally had 390,000 ultimate owner records, each representing the top of a complex enterprise. Most were the simplest form of complex enterprise, i.e., the ultimate owner was a headquarters with one or more branches under it. However, some enterprises, such as IIT, had as many as 1200 associated establishments. While relatively few firms had multi-level, complex structures, the greatest complexity was in the largest businesses; therefore, their proper treatment was essential to achieving accuracy in the data.

Due to the coding scheme used in the DMI File, parent records could only be identified as such if they were not also headquarters or subsidiaries. However, as a result of a Dun and Bradstreet editing error, these recognizable non-headquarters, parent records had had their firm employment figures replaced with establishment employment data, leaving us no indication of the real size of the firm. In order to check for internal consistency and to summarize the data for each complex enterprise, our analysis of complex firms had to work from the bottom up to the ultimate owner.

IV. Correction of Inconsistencies within Complex Establishments

Before reorganizing the complex establishment records into family groups for each firm, it was necessary to ensure logical consistency among the organizational indicators and pointers within each establishment record. Nearly 100,000 records showed evidence of incomplete or conflicting indicators and pointers. We analyzed the sources of logical inconsistencies in the complex establishment records and devised conservative correction procedures. In different cases, conservative meant either minimum change, minimum loss of data, or minimum difficulty in future processing.

We initially identified three distinct sources of inconsistencies:

1. Investigator errors in specifying indicator codes or pointers to parents and headquarters, or keypunch errors in transcribing these data.

2. Time lags between the updating of establishment data on the DMI file and the updating of ultimate owner pointers derived from the Duns semi-annual company affiliation update procedure.

3. Past errors in computer programming or operation that were either undetected or uncorrected. The effects of these errors sometimes interacted to obscure the primary problem.

Algorithms which detected and corrected fifteen types of inconsistencies were used to make organizational pointers consistent with the organizational status codes. Each branch record was required to have legitimate headquarters and ultimate owner pointers (legitimate defined as different from self). About 39,000 records were coded as branches and had legitimate headquarters pointers, but either pointed to themselves or to non-existent records as their ultimate owner. The pointers were corrected for most of these branches. The remainder, which lacked sufficient information to permit correction, were converted to single, non-branches. Similarly, each subsidiary was required to have a legitimate parent pointer. Most inconsistencies at this level were between the parent and ultimate owner pointers.

Establishments which were neither branches nor subsidiaries, but which had ultimate owner pointers, were required to point to themselves as ultimate owners.

V. Analysis and Correction of Enterprise Structure

After the corrections had been applied to achieve internal consistency in the codes and pointers, the establishments on the file were grouped into enterprises. Establishment records were ordered by ultimate owner; branches and subsidiaries were grouped together within each enterprise. The file was then subjected to completeness analysis. The first step was to identify and extract "topless" enterprises on the file. A "topless" enterprise was one with no establishment record on the DMI File corresponding to the ultimate owner pointers in the member establishments. A designated top or ultimate owner would not have been grouped with its subsidiaries and branches if it were, in fact, a subsidiary of another establishment. Erroneous or incomplete reporting in the family members could produce a misplacement of this type. A search was made throughout the tree file for the reported owners, which were found for about 11,000 of the apparently topless enterprises. Their ultimate owner field was corrected to indicate the actual ultimate owner instead of the establishment mistakenly reported to be the ultimate owner.

About 12,000 establishment records in topless enterprises still lacked ultimate owners. Of these, 7,400 establishment records were coded as branches or subsidiaries, but were related to no other record found on the file. These establishment records were grouped by major industrial group (two-digit SIC) and an ultimate owner record was imputed for each of the 72 groups. About 4,500 establishments remained in 891 topless multi-member enterprises. An ultimate owner record was created for each of these topless enterprises. The SIC code assigned to the imputed ultimate owner record was that of the major industry group accounting for the most employment in the family. All the enterprise family groups then had a top establishment; reported, corrected, or imputed.

The second step of the completeness analysis comprised the examination of each enterprise structure and the verification of its ownership linkages. The analysis was done from the bottom up - first for each subsidiary, then for the top level of the enterprise. A check was made to ensure that all establishments pointed to by the branches in the family were present and were marked as headquarters. Each subsidiary was checked to verify that its immediate parent was either another subsidiary in the same family or the top of the family. Any headquarters which had no branches pointing to it and had total employment equal to or lower than establishment employment had its headquarter status revoked. The same rules were applied to the ultimate owner of the family, ensuring that if the top were coded as a headquarters, it had branches pointing to it. A record coded as an ultimate owner which had no branch or subsidiary records pointing to it and which had equivalent or lower total employment than establishment employment was changed into a single establishment firm. As with all single establishment firms, its total employment was set equal to its establishment employment. A total of 68,000 complex establishments were reclassified as single.

VI. Employment Adjustment and Imputation of Branches

Theoretically, employment total in a subsidiary headquarters record should represent the aggregated establishment employment of itself and all its branches. The total employment figure reported for the ultimate owner or top of an enterprise includes all employment of all domestic establishments owned by the top - subsidiaries and branches. The total employment of a non-headquarters subsidiary should

represent only that establishment's employment. When there was evidence that these principles were violated, we reconciled the inconsistencies either by adjusting the total employment figures or by imputing an additional branch establishment.

Adjustments to employment total were needed under two circumstances. The first type occurred when aggregate establishment employment was greater than reported employment total. In this situation the total employment field was reset to the sum of establishment employment. The second type of adjustment occurred when reported total employment was larger than the aggregate establishment employment, but the difference was considered too insignificant to justify the imputation of an additional branch establishment. This small difference could be due to rounding of large employment figures or to updating of employment figures for some, but not all, of the establishments in an enterprise. The employment difference was considered insignificant when any of the following were true:

- a) employment difference was two or less,
- b) employment difference was less than 10 and total employment was greater than 1,000 or
- c) employment difference was less than 100 and total employment was greater than 10,000. Insignificant differences were corrected by resetting the total employment to aggregate establishment employment.

A new branch establishment was imputed when it seemed reasonable to assume that the discrepancy in employment figures arose from the failure to report separate establishment data for some members of the enterprise. This occurred whenever employment total was significantly larger than aggregate establishment employment. Imputed branch records were given unique identifying numbers, and the state code, the ultimate DUNS number, and the headquarters DUNS number of the top record in the family (or sub-family). Their SIC code and industry division were specified as those of the rest of the firm as a whole, determined by the rules used by the Census Bureau for classifying enterprise data for County Business Patterns. Establishment employment for the new branch was set equal to the employment discrepancy between the enterprise employment and the aggregate establishment employment, so that it reconciled the two.

These principles for reconciliation of employment data by adjustment or branch imputation were applied on two levels. First, employment reporting was reconciled in subsidiary groups - that is, parts of enterprises consisting of a subsidiary headquarters and its branches. Then the reconciliation for the ultimate owner was done using the same principles on a full enterprise basis.

Application of these principles at both levels increased both the total number of establishments and the aggregate employment levels. At the subsidiary headquarters level, employment was adjusted in 13,000 records, and 18,000 branch records were imputed. Processing on the full enterprise basis, another 115,000 records had their employment adjusted, and 202,000 branch records were imputed.

VII. Refinement of Branch Imputation

The procedure for imputing a branch to each firm whose employment data indicated incomplete reporting of member establishments did not address the question of how many establishments were not reported. Indeed, we studied a sample of firms and were not able to deduce any general rule to estimate from each firm's data how many branches were missing. We did know that sales branches of manufacturing firms were not reported, but we could not generalize about how many sales branches a given firm should have. We considered relating size of imputed branches to size of reported branches for that firm, but that size often seemed ridiculously small and would have ballooned the number of branch establishments enormously.

Inspection of establishment reporting for a sample of firms showed that, as might have been expected, Dun and Bradstreet frequently reported relatively small central administrative offices, while not covering the large productive branch establishments. Thus, a firm with 200,000 employees might have reported fifty branches or subsidiaries with an average employment of two hundred. The 190,000 employees unaccounted for might very well have been in 10 branches with 19,000 employees in each location. It is certainly unlikely that it would really comprise 950 unreported branches with 200 employees each (except perhaps in retail trade or services).

Designing a reasonable scheme for breaking up the imputed branches where appropriate was essential to enhancing the data and preserving the statistical quality of the data. General rules for the refinement of the branch imputation were therefore necessary. The level of employment reported for branches differed considerably by industry division and by total firm size. Working from tabulations of average branch size by enterprise employment size class for each industry division, we estimated equations for branch size as a function of firm size. Using this calculated branch size for disaggregating imputed branches has the advantage of avoiding distortion of the reported establishment size distribution.

This calculated branch size for each firm with an imputed branch was used, not as the actual size of each imputed branch, but to determine the number of branches which should be used to represent the employment otherwise unaccounted for. The number of branches imputed for a firm was determined by dividing the firm's imputed employment figure (as represented by the employment of the single imputed branch) by the branch size calculated for that firm and rounding down to the nearest integer. Thus, no additional branch would be imputed unless the imputed employment figure was at least twice the average branch size for that size firm in that industry division. A limit of one hundred imputed branches per firm was imposed to restrict imputation for the roughly 300 large firms with most of their employment unaccounted for. Firms with total employment less than twenty were limited to a single imputed branch. This technique allocated the original 202,000 imputed branches into 428,000 branches of more appropriate size.

VIII. Linking Enterprise Data to Establishment Data

The final step in the development of enterprise data was to compute the two most commonly used enterprise characteristics — firm industry division and firm employment size class. These two descriptive data items were appended to the establishment data for each member of the firm.

For employment data, such as County Business Patterns, the Census Bureau defines the industry division of an enterprise as that industry division which accounts for the largest portion of the enterprise's payroll. Using employment as a proxy for payroll, we computed enterprise industry division for each firm by summing up establishment employment classified by the industry division of each establishment's primary SIC. The firm's industry division is the one comprising the largest portion of employment. The mining industry provides a good example of the impact of differing definitions of firm industry. If firm industry division were defined as the industry division of the top establishment of the firm, mining would have included 25,331 enterprises, which own 39,885 establishments with 1,808,000 employees. When the Census Bureau's employment based definition was used, many of these enterprises were reclassified as manufacturing, especially the large oil companies whose refining and petrochemical businesses dominate their employment. Additionally, many small enterprises primarily engaged in mining whose tops were in other industries were shifted into this industry division by the application of this definition. A small net increase in the number of enterprises and establishments resulted, but it was accompanied by a dramatic

decrease in employment. The Census Bureau's definition gives us 25,396 enterprises with 40,043 establishments and 1,035,000 employees. 3/

Another important enterprise variable is the enterprise employment size class, which we call firm size class. This is simply a coded variable representing the total employment of the firm, which, after completion of the Tree file, is the actual sum of establishment employment in all the member establishments. Having this datum in the record of each establishment belonging to a complex family permits us finally to analyze, easily and efficiently, our entire file of establishment data classified by firm size. It is usually this size class which is relevant to policy analysis. Because we have the completed Tree file and its associated establishment file, we can now compute other enterprise characteristics that might be needed for special analysis.

IX. Special Uses of Establishment Data with Associated Firm Data

The association of accurate firm size and firm industry division data with each establishment record on the 1977 USEEM provides a solid basic data set on American business establishments and firms. Ideally, this procedure would be repeated for data from several other years and a longitudinal file developed. We have already begun work on data for the 1979 USEEM.

The basic data set now available is a unique tool for analysis and for interpolation of data from other sources. Because our establishment population is well defined and is placed in the context of the owning enterprise for each establishment, USEEM provides a basis for comparison of otherwise non-comparable statistics and a framework for disaggregation of aggregated data.

Consider the question of determining the share of small business in total U.S. business, in regional business, or in particular industries. Rather detailed data on employment by employment size class are available from several sources which would be useful for looking at this question. However, most of these sources provide only establishment basis data and have various limitations on their population coverage. For each source, the comparable population in the USEEM can be defined, and factors can be calculated to convert establishment distributions to enterprise distributions, at whatever level of detail is desired.

For instance, special tabulations of 1972 and 1977 Unemployment Insurance data on employment by employment size class by industry were produced for the Small Business Administration (SBA). These tabulations show that net establishment employment growth from 1972 to 1977 in all industries except government and agriculture can be accounted for as follows:

Table 2

| 1972-1977 Net Growth in Establishment Employment (Employment in Thousands) | | |
|--|----------------------|----------------------------|
| Employment Size Class | Employment Growth | Percent of Total Growth |
| 1 - 99 | 3,807 | 51.7 |
| 100 - 999 | 2,604 | 35.4 |
| 1000 + | 945 | 12.9 |
| | 7,356 | 100.0 |

Source: Special tabulation of unpublished Unemployment Insurance data prepared for Office of Advocacy of the SBA in 1980.

The reporting unit for UI is usually an establishment, but we are interested in employment growth distributed by firm size, not establishment size. For this paper's definition of small business, firm employment under 100, we can use data from the USEEM to convert this distribution into one of small versus large firm size.

Table 3 shows the distribution of employment in the USEEM for establishments with fewer than 100 employees by establishment employment size class and by firm employment size class. Establishments that have fewer than 100 employees, which belong to enterprises with more than 100 employees we call pseudo-small. The percentage of small establishments that are actually pseudo-small is surprisingly high for the establishments with between 20 and 100 employees. In Table 3 notice that 32 percent of employment in small establishments was actually in large firms. If we assume that this employment distribution had not changed substantially since 1972 and that the growth rate for small establishments was independent of their ownership, then we can apportion UI's 3.8 million small establishment employment growth by firm size. Thus 32 percent of the growth in small business employment, 1.2 million employees, is attributable to large firms and the remaining 2.6 million is attributable to small firms. The distribution of net growth in employment shown by UI data becomes 35 percent in small firms versus 65 percent in large firms. Similar procedures can be developed to transform other establishment based distributions at any level of disaggregation, for any definition of small business.

The proportion of small establishments which are pseudo-small differs considerably in various industries. Taking one hundred employees again as the upper limit for small business size, Table 4 shows this variation for the nine industry divisions. The three industries whose small establishments are most dominated by large firms are mining (including petroleum industries); transportation, communications and public utilities; and finance, insurance and real estate. In these industries about 20 percent of the small establishments, with nearly 50 percent of the employment in all small establishments, are owned by large firms. Large firms account for about 30 percent of the small establishment employment in manufacturing, in wholesale and retail trade and in services. Even in the industries with the lowest proportions of pseudo-small (under four percent), i.e., construction and agriculture, forestries and fisheries, a substantial amount of small establishment employment (14 percent and 17 percent respectively) is controlled by large firms. Any attempt to analyze economic behavior of firms using establishment employment data should take into account these differences.

X. Summary

Completing the establishment-enterprise association in the 1977 USEEM has provided a unique resource for economic research on U.S. business. The comprehensive population coverage of the original DMI File made the effort and cost of correcting errors and inconsistencies and the reconciling of the employment data worthwhile. The procedures outlined in this paper, while having little apparent effect on the aggregate data for firms and their employment, significantly

improved the quality of establishment data. We corrected the codes and pointers for over 200,000 complex establishments. This enabled us to identify 12,000 establishments with apparently non-existent owning firms and to create imputed tops to represent their ultimate owners. The reported firm employment figure was corrected for 195,000 top and subsidiary establishments of complex firms. Finally, we imputed over 420,000 branches to 200,000 firms to compensate for the 19 million establishment employees not accounted for.

Now that the problems have been defined and the solutions tested, the process of editing DMI files from other years and reconciling their establishment and firm employment will be considerably easier. Processing of data from other years is necessary for the next level of research data development - longitudinal establishment and enterprise data files which can be used to study business births and failures, divestitures and acquisitions, and enterprise employment changes.

Table 3
Establishments with Fewer Than 100 Employees
by Establishment and Firm Employment Size Class

| Number Employees | Pseudo-Small Firm > 100 Employees | True Small Firm < 100 Employees | Total Small | Pseudo-Small/Total Small (Percent) |
|------------------|-----------------------------------|---------------------------------|-------------|------------------------------------|
| 0-4 | 71,000 | 2,402,800 | 2,473,900 | 2.9 |
| 5-9 | 67,500 | 854,000 | 921,500 | 7.3 |
| 10-19 | 132,700 | 477,800 | 610,500 | 21.7 |
| 20-49 | 170,600 | 251,900 | 422,500 | 40.4 |
| 50-99 | 82,500 | 61,200 | 143,700 | 57.4 |
| TOTAL | 524,300 | 4,047,800 | 4,572,100 | 11.5 |

Employment in Establishments
by Establishment and Firm Employment Size Class
(Employment in Thousands)

| Number Employees | Pseudo-Small Firm > 100 Employees | True Small Firm < 100 Employees | Total Small | Pseudo-Small/Total Small (Percent) |
|------------------|-----------------------------------|---------------------------------|-------------|------------------------------------|
| 0-4 | 201 | 5,716 | 5,916 | 3.4 |
| 5-9 | 450 | 5,442 | 5,892 | 7.6 |
| 10-19 | 1,900 | 6,174 | 8,074 | 23.5 |
| 20-49 | 5,168 | 7,114 | 12,282 | 42.1 |
| 50-99 | 5,532 | 3,889 | 9,422 | 58.7 |
| TOTAL | 13,250 | 28,335 | 41,585 | 31.9 |

SOURCE: Version I of USEEM (V4DMJ, Table 35).

NOTE: All establishment counts are rounded to nearest hundred. All employment figures are rounded to nearest thousand.

Table 4

FOOTNOTES

Establishments with Fewer Than 100 Employees
by Firm Employment Size Class and Industry Division

| Industry Division | Pseudo-Small Firm > 100 Employees | True Small Firm < 100 Employees | Total Small | Pseudo-Small/Total Small (Percent) |
|-------------------|-----------------------------------|---------------------------------|-------------|------------------------------------|
| FIRE * | 95,100 | 289,900 | 394,100 | 24.1 |
| MINING | 8,200 | 29,400 | 37,600 | 21.8 |
| TCPU * | 39,200 | 144,900 | 184,200 | 21.3 |
| WHLSE TRADE | 72,000 | 428,900 | 500,900 | 14.4 |
| MNFG | 57,100 | 345,800 | 402,900 | 14.2 |
| SERVICES | 92,000 | 849,800 | 941,800 | 9.8 |
| RET TRADE | 140,600 | 1,292,000 | 1,432,600 | 9.8 |
| AGRIC | 3,700 | 103,800 | 107,600 | 3.4 |
| CONSTR | 16,200 | 554,200 | 570,400 | 2.8 |
| TOTAL | 524,300 | 4,047,800 | 4,572,100 | 11.5 |

Employment of Establishments with Fewer than 100 Employees
by Firm Size and Industry Division
(Employment in Thousands)

| Industry Division | Pseudo-Small Firm > 100 Employees | True Small Firm < 100 Employees | Total Small | Pseudo-Small/Total Small (Percent) |
|-------------------|-----------------------------------|---------------------------------|-------------|------------------------------------|
| FIRE * | 1,827 | 2,014 | 3,841 | 47.6 |
| MINING | 238 | 247 | 485 | 49.1 |
| TCPU * | 1,123 | 1,164 | 2,287 | 49.1 |
| WHLSE TRADE | 1,222 | 3,024 | 4,246 | 28.8 |
| MNFG | 1,969 | 4,270 | 6,239 | 31.6 |
| SERVICES | 2,945 | 5,854 | 8,799 | 33.5 |
| RET TRADE | 3,327 | 8,085 | 11,412 | 29.2 |
| AGRIC | 111 | 550 | 661 | 16.8 |
| CONSTR | 489 | 3,128 | 3,617 | 13.5 |
| TOTAL | 13,250 | 28,335 | 41,585 | 31.9 |

* FIRE = Finance, Insurance and Real Estate

* TCPU = Transportation, Communications, Public Utilities

SOURCE: Version I of USEEM (V4DMU, Table 35).

NOTE: All establishment counts are rounded to the nearest hundred. All employment figures are rounded to nearest thousand.

1. See U.S. Establishment and Enterprise Microdata (USEEM) - Version I: File Description and also Constance Mitchell, "Employment Imputation from County Business Patterns: Methodology and Production Statistics," Working Paper No. 3. Both are available from the Small Business Microdata Project, Economic Studies Program, The Brookings Institution, Washington, D.C.

2. Approximately 85 percent of the firms on the file have gross receipts data. Associated data from other Dun and Bradstreet data files have been linked to the DMI File data in USEEM to provide data on five-year growth of sales and employment for about 24 percent of the firms. Data from D & B's Financial Statement File have been linked with 22 percent of the USEEM firms, making up an associated file that contains eleven balance sheet and income statement variables and up to five years of historical data for sales, profits and net worth.

3. Figures are derived from tabulations of the data in Version I of the interim file: TREE.V3, TREE.V5, and DMISUM.