Economic Quarterly—Volume 95, Number 4—Fall 2009—Pages 419–454

The U.S. Establishment-Size Distribution: Secular Changes and Sectoral Decomposition

Samuel E. Henly and Juan M. Sánchez

Establishment heterogeneity has been modeled in economics at least since the seminal work of Lucas (1978). More recently, this feature has been incorporated into calibrated models to provide quantitative evaluations of different mechanisms. This article aims to contribute to this literature by providing a set of facts about the establishment-size distribution since the 1970s that may be used to calibrate and test the predictions of these models.

First, this article analyzes establishment data from 1974–2006.[1] During this period, the number of workers (size) of a "representative establishment" is relatively constant. Next, the analysis turns to the dispersion of establishment sizes. The size distribution of establishments has become slightly more even. The same analysis is then applied at the sector level. Service establishments became larger and service labor became more concentrated in large establishments while opposite trends were observed in manufactures. Although these *intrasector* shifts played an important role in explaining aggregate movements, *intersector* changes were also found to be important. Finally, this article considers whether trends in the firm-size distribution resemble those

We gratefully acknowledge comments from Anne Stilwell, Devin Reilly, Kartik Athreya, Marianna Kudlyak, and Ned Prescott. All remaining errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: sam.henly@rich.frb.org; juan.m.sanchez@rich.frb.org.

[1] Two alternative production units will be considered—firms and establishments. A firm may be a collection of establishments. For instance, Walmart is one firm but it has more than 4,000 establishments in the United States.

found in establishments. They are similar, although labor became slightly more concentrated in large firms.

Davis and Haltiwanger (1989) also analyze secular trends at the establishment level.[2] In particular, they study changes in the establishment-size distribution during the period 1962–1985. First, they study how workers are distributed across establishments; they find that the "representative" worker was working in a larger establishment in 1962 than in 1985. Second, they consider the establishment-size distribution; conversely, they find that the "representative" establishment was smaller in 1962 than in 1985.[3] The opposite behavior of these series reveals a decline in the dispersion of establishment size. Davis and Haltiwanger also decompose these changes by sector. They find that "changes in the industry distribution of employment and movements in the employee size distribution within the average two-digit industry make roughly equal contributions to the secular shift towards mid-size establishments in the aggregate economy." This article extends part of their work through 2006 and complements it with an analysis of firm data and alternative statistics, figures, and decompositions. The earlier change in the first moments contrasts with the finding in this article, while the downward trend in the dispersion of establishment size continued after 1985.

Buera and Kaboski (2008) also study the evolution of the scale of production and sectoral reallocation. They emphasize the difference between the size distribution for manufactures and services establishments. Additionally, they present evidence of the rise in the size of service establishments and the reallocation of resources from manufacturing to services.[4] Our article extends their analysis by studying changes in the size distribution of manufacturing and service establishments over time.

Several studies take an interest in which distribution best fits the firm-size distribution. Gibrat (1931) finds that the log-normal distribution effectively described French industrial firms. This distribution is a consequence of the "law of proportional effect," also known as Gibrat's Law, whereby firm growth is treated as a random process and growth rates are independent of firm size (Sutton 1997). As noticed by Axtell (2001), census data display monotonically increasing numbers of progressively smaller firms, a shape the log-normal distribution cannot reproduce. Using data from the U.S. Census Bureau from 1988–1997, Axtell (2001) shows that firm size is approximately Zipf-distributed. Although we find that the aggregate distribution is relatively

[2] See also Davis and Haltiwanger (1990) and Davis, Haltiwanger, and Schuh (1996).

[3] Our article also considers the distribution of employees by establishment size and the distribution of establishments by size. Notice that while the latter describes which proportion of the establishments is of a given size, the former studies which proportion of employees work in an establishment of a given size.

[4] They also show evidence of sectoral reallocation for 30 countries.

stable, results for manufacturing and services suggest that it would be interesting to extend Axtell's analysis to the sectoral level.

Recent articles use establishments data to study economic development. They argue that the misallocation of resources among heterogeneous establishments may be a key determinant of cross-country income differences. Banerjee and Duflo (2005) conclude that "the microeconomic evidence indeed suggests that there are some sources of misallocation of capital, including credit constraints, institutional failures, and others." Restuccia and Rogerson (2008) illustrate this mechanism using a model with establishment heterogeneity similar to Hopenhayn and Rogerson (1993). In a similar framework, Hsieh and Klenow (2007) find that productivity would increase by 30–50 percent in China and 40–60 percent in India "if capital and labor were reallocated to equalize marginal products across plants to the extent observed in the U.S." Similarly, Greenwood, Sánchez, and Wang (2008) study the role of informational frictions for economic development in a model with establishments heterogeneity.[5] All the theories above analyze mechanisms that may contribute to an explanation of differences in income across countries. The calibrations of these and similar models generally use targets from the size distribution. For instance, Restuccia and Rogerson (2008) use the 2000 establishment size distribution and Greenwood, Sánchez, and Wang (2008) use the Lorenz curve for the distribution of employment by establishment size for 1974. The subsequent sections of this article present evidence for size distributions of establishments and firms and supply a set of stylized facts that new theories in this strand of literature may find useful as calibration targets. Perhaps more importantly, these sections analyze secular changes in the size distribution that could be used to test the predictions of these models. For example, we find that the average size of establishments is fairly constant (or slightly decreasing) over the last 30 years. This finding supports models in which the average size is constant on the balanced-growth path.

The remainder of the article is organized as follows. Section 1 introduces and summarizes our findings. Section 2 describes the secular changes in the establishment-size distribution. The decomposition of secular changes into changes in the sectoral composition (intersector) and distribution changes within each sector (intrasector) is undertaken in Section 3. A description of the data on firms, as an alternative to establishments, is presented in Section 4. Finally, Section 5 concludes. An Appendix presents detailed information about data sources, formulae used to compute the statistics, and some figures and tables.

---

[5] See also Caselli and Gennaioli (2003); Amaral and Quintin (2007); Alfaro, Charlton, and Kanczuk (2008); Bartelsman, Haltiwanger, and Scarpetta (2008); Buera and Shin (2008); Guner, Ventura, and Yi (2008); and Castro, Clementi, and McDonald (2009).

1. PRODUCTION UNIT SIZE TRENDS, 1970–2006

In the sections below, several statistics are defined and used to evaluate the distributions of productive units and their workers from the 1970s to 2006. The aggregate economy, as well as two component sectors (manufacturing and services), are considered in each analysis.

Section 2 develops statistics and functions that are used in the analysis of trends in establishment size and shifts in the dispersion of establishments and workers. We find that the aggregate establishment size changes negligibly. Manufacturing establishments are very large and shrink over time, while service establishments are initially smaller than average but become much larger by 2006. Variation of establishment size does not change significantly apart from a small increase in the service sector. The distribution of employees across establishments becomes slightly more even. This trend is driven by the decline of large manufacturing firms and dampened by increased labor concentration in services.

Section 3 decomposes, by sector, several statistics introduced in Section 2. The results are used to disentangle changes in aggregate statistics caused by intrasector distribution movements from those caused by shifts in the sectoral composition of the aggregate (intersector changes). We find that both intra- and intersector movements are important, but the importance of each varies by statistic.

Section 4 examines the question of whether and when firm distribution patterns should resemble those found in establishments. We argue that movements in establishment distributions should be more similar to those in firms when large firms are composed of relatively large establishments, and present evidence is consistent with this hypothesis. Trends in the aggregate and sectoral distributions of firms and employees across firms generally conform to trends at the establishment level.

2. SECULAR CHANGES IN THE SIZE DISTRIBUTION OF ESTABLISHMENTS

The U.S. Census Bureau (USCB) publishes annual data on establishments in their County Business Patterns series. This section presents a variety of statistics derived from these data. The statistics describe the size distribution of establishments and the dispersion of labor and establishments across establishments. Major trends in these statistics since 1974 are noted and depicted in Figures 1–8.

County Business Patterns Data

County Business Patterns (CBP), released by the USCB annually since 1964, contains tables listing establishment quantity, worker quantity, and payroll by

Table 1 Example Establishment Data

Size Group	Establishment Size	Number of Establishments
1–2 Workers (Small)	1	5
	2	2
3–4 Workers (Large)	3	2
	4	1

establishment size groups. For example, CBP tables in any given year list the number of establishments employing 20–49 workers, the number of people employed by those establishments, and other data (like payroll) not used in this article. Similar data are provided for other establishment size groups (1–4 workers, 5–9 workers, etc.). This information is given for the aggregate and also by SIC (1997 and earlier) or NAICS (1998 onward) industry category. We use data for years 1974 and later due to a significant methodological shift taking place between 1973 and 1974.[6]

A caveat is in order. In the service sector, data for years before and after 1997 are not directly comparable: After 1997, an establishment's sector was determined by the North American Industrial Classification System (NAICS), which is not easily reconciled with the Standard Industrial Classification (SIC) system used for the same purpose in previous years.[7] Consequently, analysis of labor concentration across service sector establishments treats SIC years (1974–1997) and NAICS years (1998–2006) separately. The composition of the manufacturing sector also changes with NAICS, but a single series is available under each system and differences are minimal.

Mean Establishment Size and Coworker Mean Size

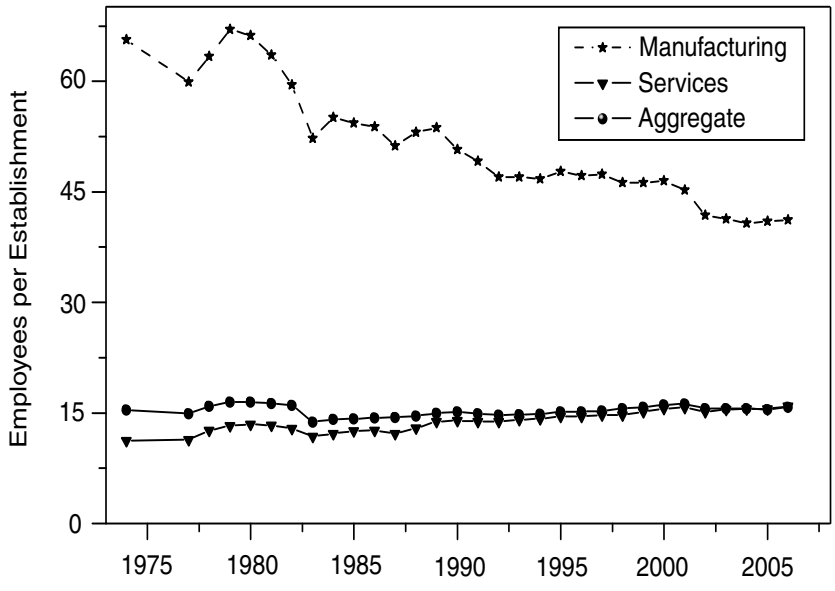
Two different measures of mean size will be considered to describe the size of a "representative" establishment. Given data restrictions, the comparison of these two measures will be used later to study the dispersion of establishments by size.

It may be useful to consider the world described in Table 1, where establishments have between one and four employees (inclusive) and are separated into two size groups: two or fewer workers and three or more workers. In

[6] Some data were retrieved from the National Historical Geographic Information System, an online database operated by the Minnesota Population Center (Ruggles et al. 2009).

[7] Under the SIC system, a single series summing all portions of a "service" sector was available. NAICS split the sector into numerous constituents (educational services; health care and social assistance; professional, technical, and scientific services; and so on). A composite service sector was constructed from these NAICS service subsectors (see Appendix) but it was not possible to precisely recreate the SIC service sector's composition.

Figure 1 Mean Size of Establishments, 1974–2006



this world, a "small" group is comprised of seven establishments employing a total of nine workers; the remaining three establishments form a "large" group employing 10 workers.

Establishment mean size

We begin by asking: What is the average establishment size across establishments? The answer is the mean of the distribution of establishments by establishment size, referred to hereafter as the *mean size of establishments* (or simply as the *establishment mean*) and denoted $E(esize)$. Denote index establishment size groups by i. Then, we obtain the establishment mean by taking a weighted sum of the expected size of establishments within each size group i:

$$E(esize) = \sum_i E(esize \mid egroup = i) * P(egroup = i). \qquad (1)$$

Here, $egroup = i$ is the condition in which an establishment is a member of size group i.[8] Considering our example world, we find that

$$E(esize) = [9/7] * (7/10) + [10/3] * (3/10) = 1.9. \qquad (2)$$

[8] Calculations of expected values and probabilities are detailed in the Appendix.

Figure 1 displays the mean size of establishments between 1974 and 2006. Across the period, this mean changes negligibly: In 1974, the average establishment employed about 15 workers, a figure that ranged between 14 and 16 workers in subsequent years through 2006. This constancy in the aggregate masks significant shifts at the sector level. The average manufacturing establishment size fell from almost 70 employees in the late 1970s to about 41 employees in 2006. The greatest decline occurred between 1979 and 1983, when the average size dropped from 67 employees to 52 employees. In spite of this decline, manufacturing establishments tend to be much larger than other establishments in all years. For instance, in 1974 the average manufacturing establishment employed about 50 more workers than the aggregate economy's average establishment; this gap was halved by 2006. Contemporaneously, the average service sector establishment increased in size, from about 11 workers in 1974 to 14.7 workers in 1997 and from 14.8 workers in 1998 to 16 workers in 2006.

Coworker mean size

What is the average number of coworkers across workers? The answer is the mean of the distribution of workers by establishment size, referred to hereafter as the *coworker mean size of establishments* or simply the *coworker mean*, denoted $E(wsize)$. This statistic is interesting because it may vary even when the mean size of establishments is constant.[9] The following formula can be used to compute this measure:

$$E(wsize) = \sum_i E(wsize \mid wgroup = i) * P(wgroup = i), \qquad (3)$$

where $wgroup = i$ denotes a worker who is employed by an establishment in size group i. In our example, we have data that allow us to compute $E(wsize)$ directly:

$$\begin{aligned} E(wsize) &= [\frac{((1*5)+(2*4))}{9}] * (\frac{9}{19}) + [\frac{((3*6)+(4*4))}{10}] * (\frac{10}{19}) \\ &\approx 2.47. \end{aligned} \qquad (4)$$

Unfortunately, $E(wsize)$ cannot be computed directly from public CBP data because we are unable to obtain $E(wsize \mid wgroup = i)$ without information about the distribution of workers within size groups. We use an alternative method of computation that employs an assumption about the distribution of establishments within size groups.[10]

[9] This was actually the case for the time period studied by Davis and Haltiwanger (1989).
[10] See details in the Appendix.

Figure 2 Coworker Establishments Mean Size, 1974–2006

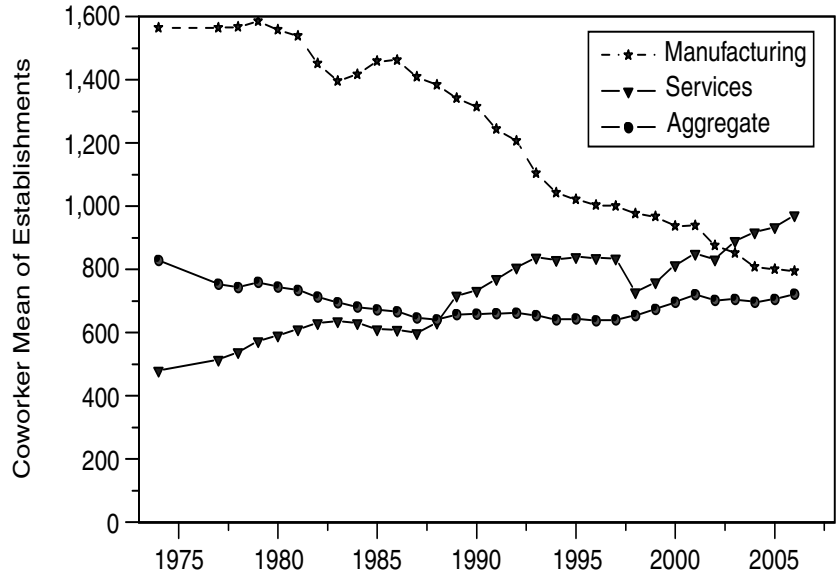


Figure 2 shows the coworker mean size of establishments. As expected, worker mean size is much greater than establishment mean size. In 1974, the worker mean stands around 830 at the aggregate level, 1,560 for manufactures, and 480 for services. Subsequent trends resemble those for the mean size of establishments. The aggregate worker mean remains fairly flat through 2006, dropping 11 percent. Simultaneously, the coworker mean in manufactures is halved (falling from 1,560 to 760) even as the services coworker mean doubles (480 to 970).

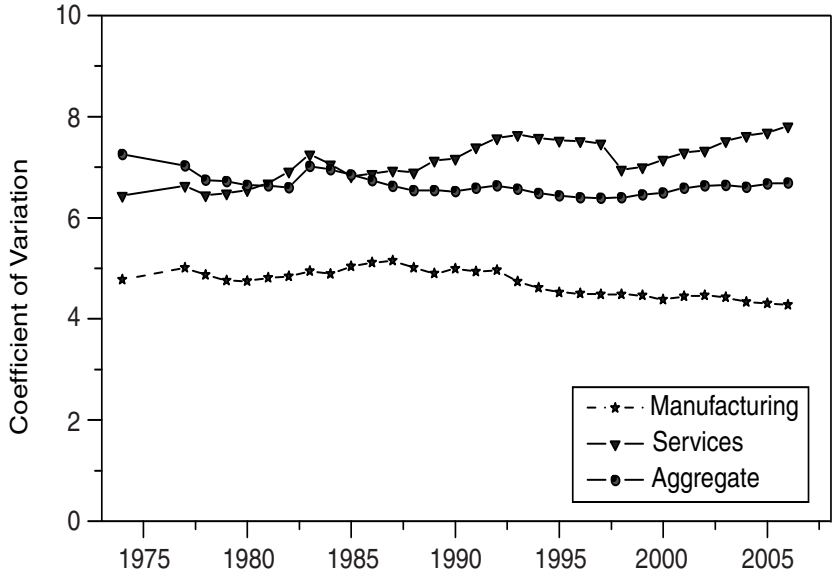
Establishment Size Dispersion and Employment Concentration

Coefficient of variation

The statistic used to analyze the dispersion of establishment size is the coefficient of variation (CV). It measures the dispersion of establishment size relative to the mean size.[11]

[11] This statistic is computed from equation (18) in the Appendix.

Figure 3 Coefficient of Variation of Establishments Size, 1974–2006

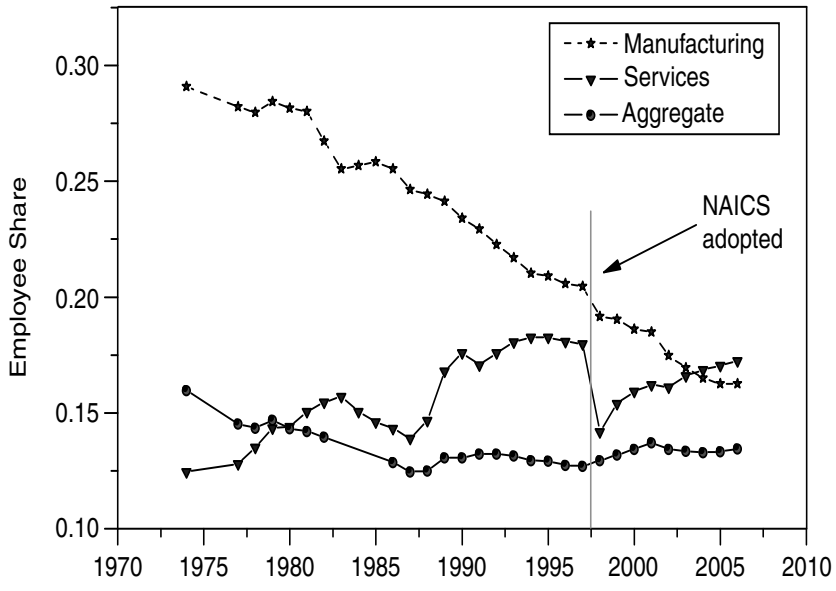


The coefficients of variation for the aggregate and for industries are displayed in Figure 3. In the aggregate, this measure fell about 8 percent from 1974 to 2006 (7.2 to 6.1). The coefficient also fell slightly in the manufacturing sector, from 4.7 to 4.2; note that this figure indicates a much lower variation in establishment size than is present in services or the aggregate. Service establishments actually saw their coefficient increase about 21 percent (6.3 to 7.8).

Large establishment employment share

The fraction of workers employed by very large establishments (those with more than 1,000 workers) serves as a simple measure of labor concentration (Figure 4). In the aggregate this figure decreased slightly. Very large establishments employed about 16 percent of all workers in 1974. By 2006, they were responsible for only 13 percent of employment, although this number had earlier dipped to a 1987 nadir of 12.5 percent. In the manufacturing sector, a decline in large establishment employment share was observed. Large establishments employed 29 percent of manufacturing workers in 1974; in 2006, they employed only 16 percent. Finally, the large establishment share of service labor moved erratically upward. In this sector the employment share

Figure 4 Employment Share of Large Establishments, 1974–2006



increased from 12.5 percent in 1974 to about 18 percent between 1990–1997; from 1998–2006, the share increased from 14 percent to 17 percent.

Lorenz curve

One frequently employed instrument for the analysis of inequality is the Lorenz curve. This measure of the distribution of labor across establishments is independent of the absolute size of establishments. Thus, if all establishments grow or shrink proportionally, there are no changes in the Lorenz curve.

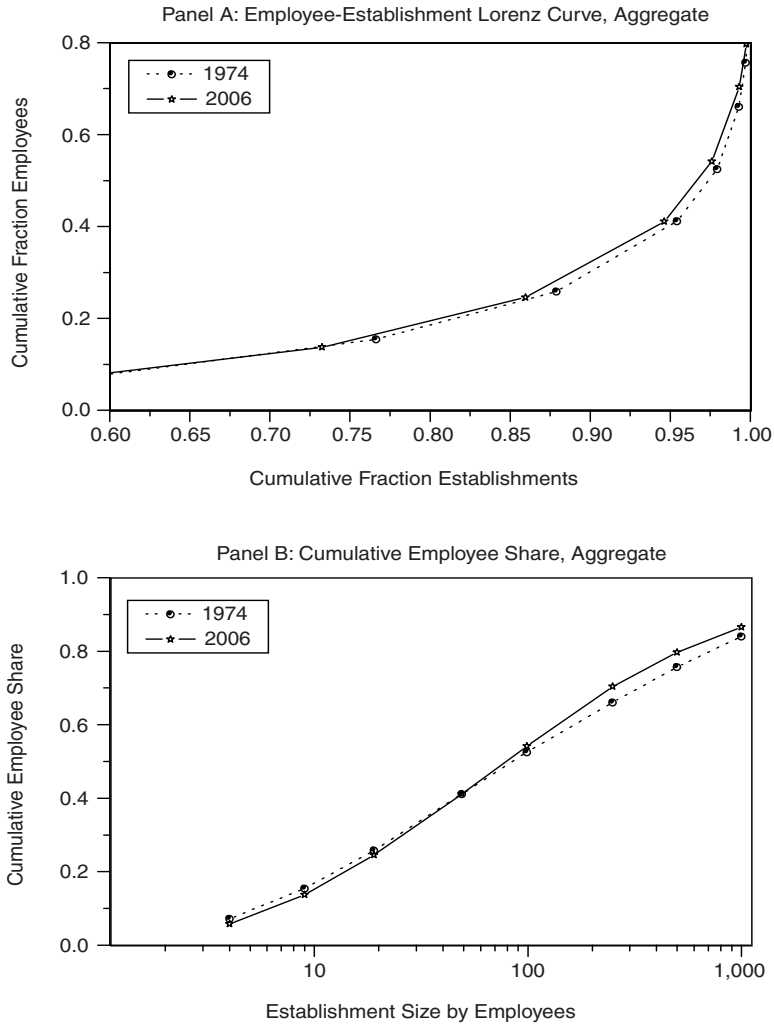
Here, a Lorenz curve represents the fraction y of total workers employed by the fraction x of total establishments employing the smallest number of workers. A 45° line means that all establishments employ the same number of workers; the further a curve is below this line, the greater the unevenness in worker distribution across establishments. Given the data restriction, we have values for the Lorenz function, L, at the upper bound of each size group i:

$$L(P(egroup \leq i)) = P(wgroup \leq i). \tag{5}$$

The function is linearly interpolated elsewhere.

Panel A of Figure 5 shows the Lorenz curve for the distribution of labor across establishments. This curve shifted slightly upward over time, suggesting a decrease in labor concentration. This movement is minor: In 1974, the largest 5 percent establishment employed about 60 percent of the country's

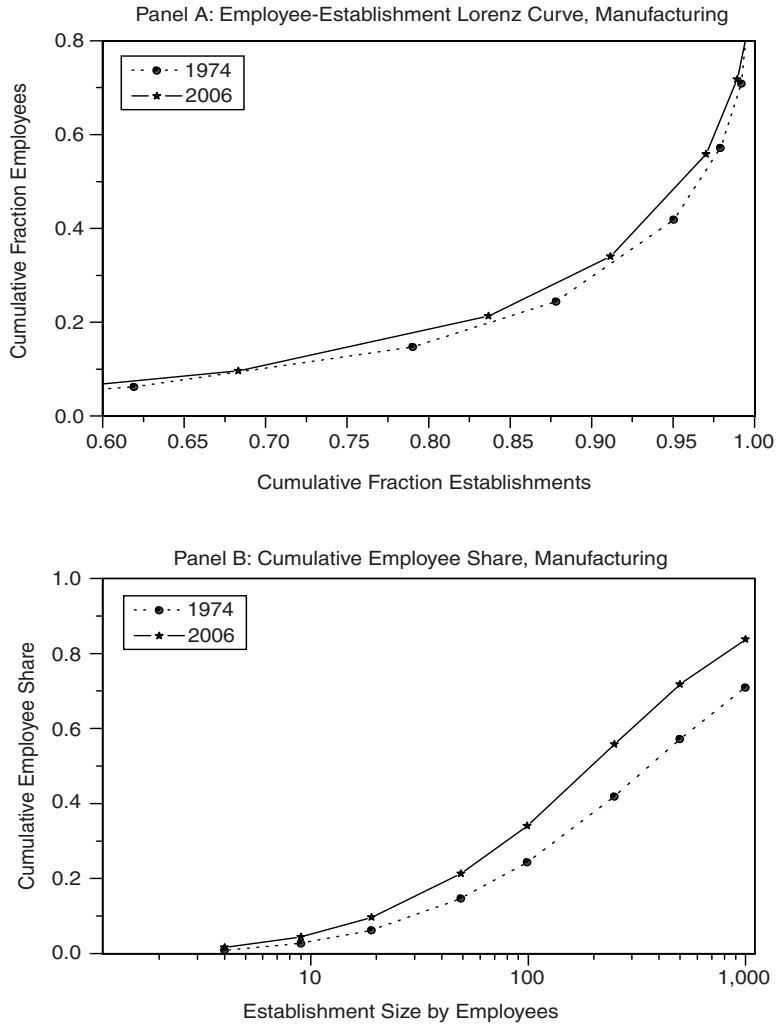
Figure 5 Establishment-Size Distribution; Aggregate Economy, 1974–2006



workers. In 2006, the same icosile employed about 57 percent of the work force.

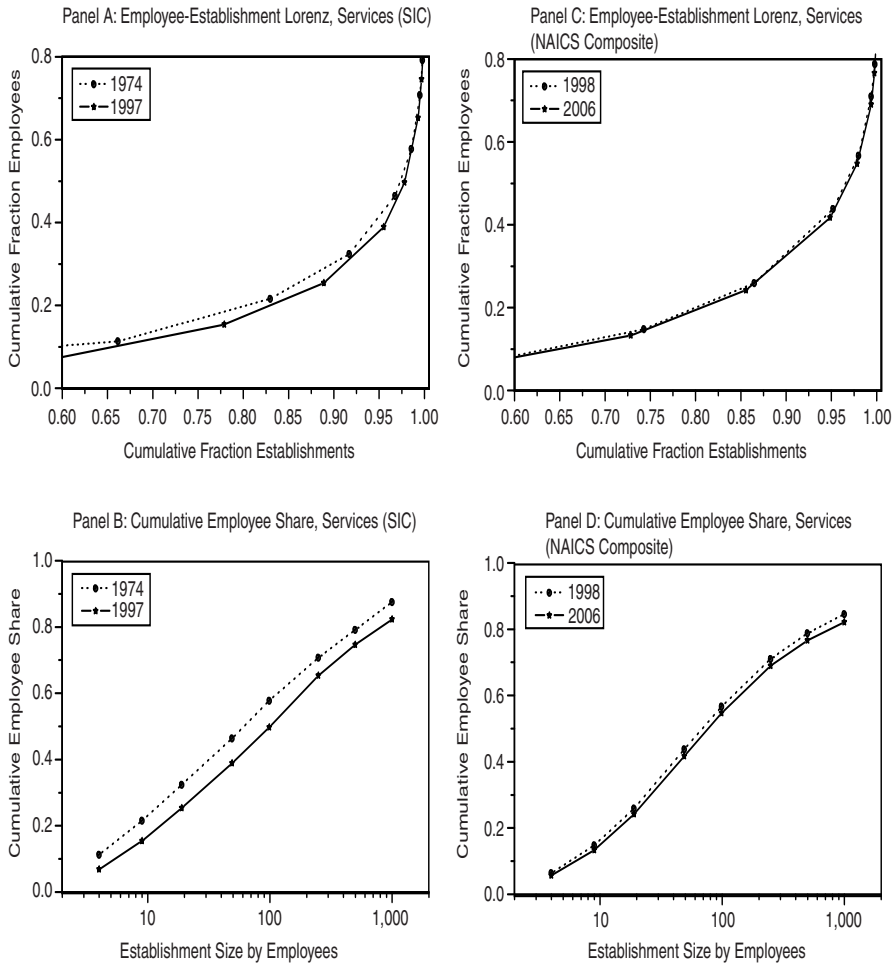
The manufacturing sector's Lorenz curve is found in Panel A of Figure 6. The curve shows a clear shift upward near the top of the scale from 1974 to 2006, as the employee share of the top 5 percent establishments fell from 58.2 percent to 51.7 percent. Workers, then, became more evenly distributed among manufacturing establishments.

Figure 6 Establishment-Size Distribution; Manufacturing Sector, 1974–2006



Service-sector Lorenz curves are located in Panels A and C of Figure 7. Over the SIC years (Panel A) the employee-establishment Lorenz curve shifted downward: The top 5 percent establishments employed about 58 percent of all service workers in 1974 and 62 percent in 1997, reflecting a greater concentration of employment in the largest service establishments. Service labor also became more concentrated in large establishments in the NAICS

Figure 7 Establishment-Size Distribution; Service Sector, 1974–2006



period (Panel C) when the largest 5 percent establishment employment share rose from 1998 (56.6 percent) to 2006 (57.6 percent).

Cumulative employee distributions

To consider the distribution of workers across establishments without explicit disregard for the absolute size of establishments (in contrast to the Lorenz curve), we construct the cumulative distribution function (CDF). This function provides the share of employment held by establishments of or less than a

particular size and is computed at the upper bound of each size group, max_i:

$$CDF(max_i) = P(wgroup \leq i), \tag{6}$$

and then linearly interpolated elsewhere.

Panel B of Figure 5 plots the CDF for the aggregate. This graph shows that the distribution of labor across establishments shifted toward mid-size firms between 1974 and 2006. In 1974, small establishments (10 or fewer employees) and larger establishments (more than 500 employees) are responsible for larger shares of total employment than in 2006. This change is visible as the 2006 curve begins below the 1974 curve but rises more quickly through the mid-size establishments. In both years, employment is nearly evenly divided between establishments with more than and fewer than 100 workers: Establishments with 99 or fewer workers employed 53 percent of the work force in 1974 and 54 percent in 2006.

The cumulative employment curve in Panel B of Figure 6 shows that every size group of manufacturing establishments below 500 workers increased its employee share from 1974 to 2006. Manufacturing establishments employing fewer than 250 workers held 56 percent of the manufacturing employment share in 2006, up from only 42 percent in 1974.

Conversely, in both SIC and NAICS periods, cumulative employment share curves for services (Figure 7, Panels B and D) moved to the right, implying a broad increase in the size of service establishments (recall data in Figures 1 and 2). Establishments employing fewer than 1,000 workers saw their employee share drop from 88 percent to 82 percent between 1974 and 1997 and from 85 percent to 82 percent between 1998 and 2006.

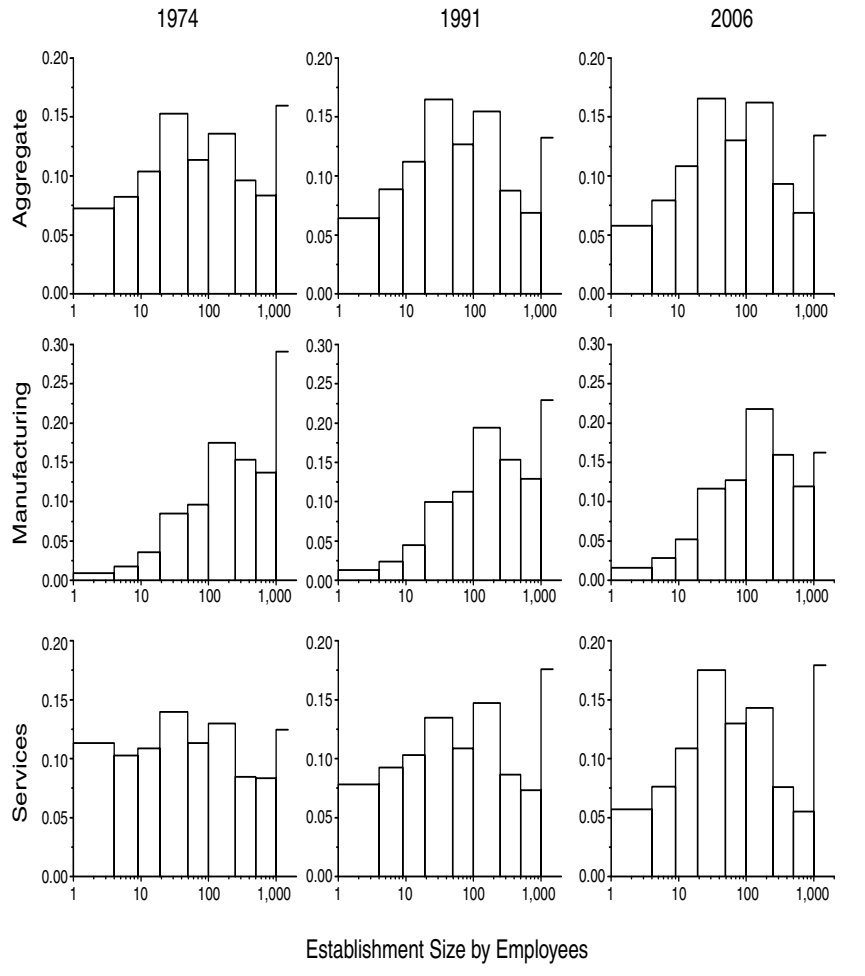
Histograms

While the CDF is useful for revealing shifts in the distribution of labor across establishments, simple histograms of the distribution of labor across establishments are helpful to identify which size groups are actually responsible for those shifts. This function is computed as

$$f((min_i - 1, max_i]) = P(wgroup = i). \tag{7}$$

where min_i and max_i are, respectively, the lower and upper establishment size bounds for size group i. The histogram for distribution of labor among size categories at the aggregate level is depicted in the top row of Figure 8. These histograms show movement of worker share from the smallest and largest establishments into establishments of intermediate size. The employee share of the smallest establishment size group decreases (1–9 workers, 15.5 percent to 13.7 percent) while intermediate size categories see their employee share increase. Establishments with 10–249 workers employed 50.6 percent of the labor force in 1974, and their share increased to 56.7 percent by 2006. Larger establishments (250–999 employees) lose employment share (18 percent to

Figure 8 Histograms for the Distribution of Labor Across Establishments



16.2 percent) as do the largest establishments (1,000 or more employees; 16 percent to 13.4 percent). Large establishments lost the most share before 1991, while small establishments lost the most after 1991.

Figure 8 also contains histograms illustrating the labor distribution across manufacturing establishments. As in previous figures, it is apparent that manufacturing sector employment was less concentrated in large establishments in 2006 than in 1974. Every establishment size group of 499 employees or fewer saw significant increases in its employment share from 1974 to 1991 and again from 1991 to 2006. Establishments employing 100–249 workers saw

the greatest increase over the entire period, employing about 17.5 percent of manufacturing workers in 1974 but 21.8 percent in 2006. By contrast, the size group 500–999 workers saw its employment share decrease from an initial 13.7 percent to 12.0 percent over the same period. This movement is in the same direction as the 13-percentage-point decline in the employment share of manufacturing establishments with more than 1,000 workers.

As noted earlier, the service sector is more difficult to probe due to differences in its composition before and after 1997. The last row of histograms in Figure 8 show that between 1974 and 1991, both years using the SIC service sector, the smallest service establishments (1–19 workers) saw their employee share drop from 32 percent to 27 percent. Intermediate size categories (20–249 workers) increased their employee share slightly, from 38 percent to 39 percent, and the largest size categories depicted (250–999 workers) lost 1 percentage point of total employee share (17 percent to 16 percent). The largest size group (1,000 or more employees) accounted for most of the balance as between 1974 and 1991 its share increased from 12 percent to about 18 percent. A histogram for 2006 shows further erosion in the employment share of the smallest and largest establishments depicted, but these data cannot be directly compared with data from 1974 or 1991.

3. SECTORAL DECOMPOSITION OF SECULAR CHANGES

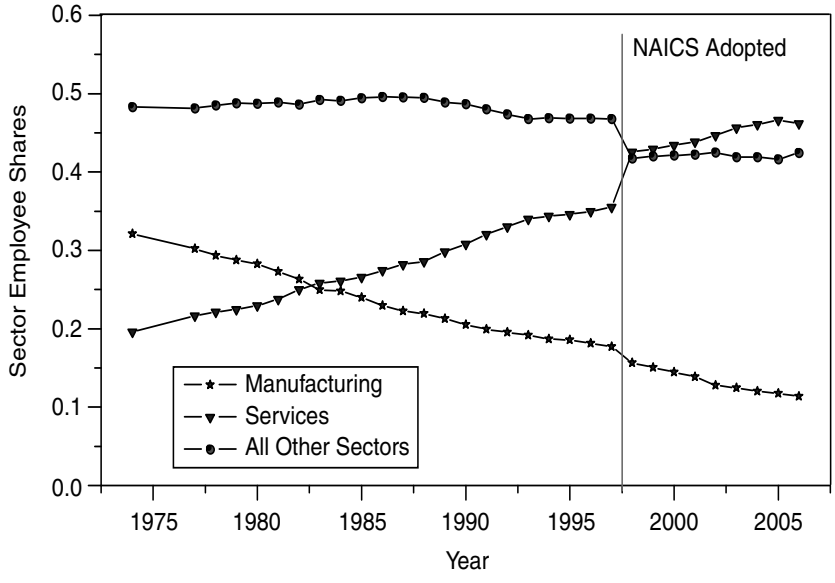
Changes in the Sectoral Composition

Previous sections demonstrated that, broadly speaking, manufacturing establishments have become smaller and service establishments have become larger since the mid-1970s. The distribution of workers became more even across manufacturing establishments and less even across service establishments. These sector level trends offset one another in the aggregate economy. However, to better understand the cause of the slight decline in overall establishment size and labor concentration, it is also necessary to consider changes in the relative share of the service and manufacturing sectors over time.

Two types of effects can be cited as contributors to observed trends in the aggregate distribution of labor across establishments. First are *intrasector* movements of labor; these are described for manufacturing and service sector establishments in the previous section. Intrasector movements of labor include shifts of employment share of different establishment size categories and changes in the dispersion of labor across establishments. The aggregate can also be affected by *intersector* forces as the relative labor and establishment share of different sectors change.

Figure 9 displays the sector shares of total employment from 1974 to 2006, and Figure 10 shows the sector share of establishments for the same period. The pattern is similar in both figures. The participation of other sectors

Figure 9 Employment Share by Sector, 1974–2006



is relatively constant,[12] only decreasing slightly in establishments; service sector participation rose and manufactures participation fell. Changes are more notable in terms of worker shares: manufacturing had 32 percent in 1974 and 11 percent in 2006, while services had 19 percent in 1974 and 46 percent in 2006. During the same period, the establishment share of manufacturing dropped from 8 percent to 4 percent while the services establishment share rose from 27 percent to 47 percent.

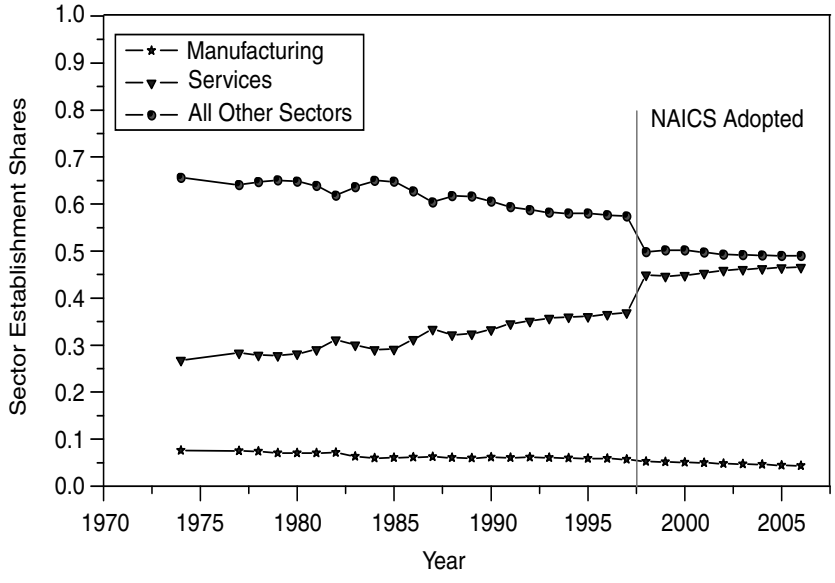
Computation

Any aggregate statistic is a weighted average of the sectoral values of that statistic. Therefore, it can be decomposed into its sectoral constituents. As an example, consider the mean size of establishments, the first statistic that will be decomposed. It can be written as

$$E(esize) = \sum_{s} E(esize \mid esector = s) * P(esector = s), \qquad (8)$$

[12] The main change seems to be in 1997, when a new sector classification system was adopted (NAICS). Of course, this implies that this change does not have economic meaning. These data were derived from County Business Patterns figures.

Figure 10 Establishment Share by Sector, 1974–2006



where s is a sector index and $esector = s$ denotes that an establishment operates in sector s. By separating services, manufacturing, and the combined other sectors, and simplifying the notation, the mean size of establishments can be written

$$E(esize) = n^{\text{serv}} E^{\text{serv}} + n^{\text{manuf}} E^{\text{manuf}} + n^{\text{other}} E^{\text{other}}, \tag{9}$$

where $E^{\text{s}} = E(esize \mid esector = s)$ and n^{s} is the establishment-share of each sector, $n^{\text{s}} = P(esector = s)$.

This decomposition may be used to answer two questions: (1) What would the value of a statistic (the establishment mean in this example) be if the intersector weights had stayed at their 1974 values? and (2) what value would the statistic have taken if the intrasector value of the statistic had stayed the same as in 1974? The first question is answered by computing a counterfactual statistic,

$$E(\widetilde{esize})_t = n^{\text{serv}}_{1974} E^{\text{serv}}_t + n^{\text{manuf}}_{1974} E^{\text{manuf}}_t + n^{\text{other}}_{1974} E^{\text{other}}_t. \tag{10}$$

Similarly, the second question is answered by computing another counterfactual statistic,

$$E(\widehat{esize})_t = n^{\text{serv}}_t E^{\text{serv}}_{1974} + n^{\text{manuf}}_t E^{\text{manuf}}_{1974} + n^{\text{other}}_t E^{\text{other}}_{1974}. \tag{11}$$

Other statistics can be decomposed in a similar manner. The only difference is that some of them require a different weight, the sector employment share,

Table 2 Sectoral Decomposition of Changes Between 1974–2006

Statistic	Aggregate Value		Manufactures Weight	Manufactures Value	Services Weight	Services Value	Other Sectors Weight	Other Sectors Value
Mean Size								
Year 1974	15.447	=	0.076	65.6	0.268	11.3	0.656	11.4
Year 2006	15.776	=	0.044	41.2	0.466	15.6	0.491	13.6
Intrasector	16.263	=	0.076	41.2	0.268	15.6	0.656	13.6
Intersector	13.690	=	0.044	65.6	0.466	11.3	0.491	11.4
Coworker Mean								
Year 1974	845.095	=	0.321	1,563.4	0.196	479.1	0.483	516.2
Year 2006	754.655	=	0.114	793.9	0.462	970.8	0.424	508.8
Intrasector	690.784	=	0.321	793.9	0.196	970.8	0.483	508.8
Intersector	618.081	=	0.114	1,563.4	0.462	479.1	0.424	516.2
Coefficients of Variation*								
Year 1974	7.329	=	0.076	102,598.9	0.268	5,400.1	0.656	5,871.8
Year 2006	6.844	=	0.044	32,687.4	0.466	15,185.9	0.491	6,944.3
Intrasector	6.401	=	0.076	32,687.4	0.268	15,185.9	0.656	6,944.3
Intersector	7.186	=	0.044	102,598.9	0.466	5,400.1	0.491	5,871.8

Notes: *Aggregate coefficients of variation are calculated here as the square root of the sum of the products of sector weights and variances, all over the mean establishment size.

defined as $e^s = P(wsector = s)$, where $wsector = s$ is the condition that a worker is employed at an establishment in sector s. Notice that e^s and n^s are the shares presented in Figures 9 and 10, respectively.

Decomposition Results

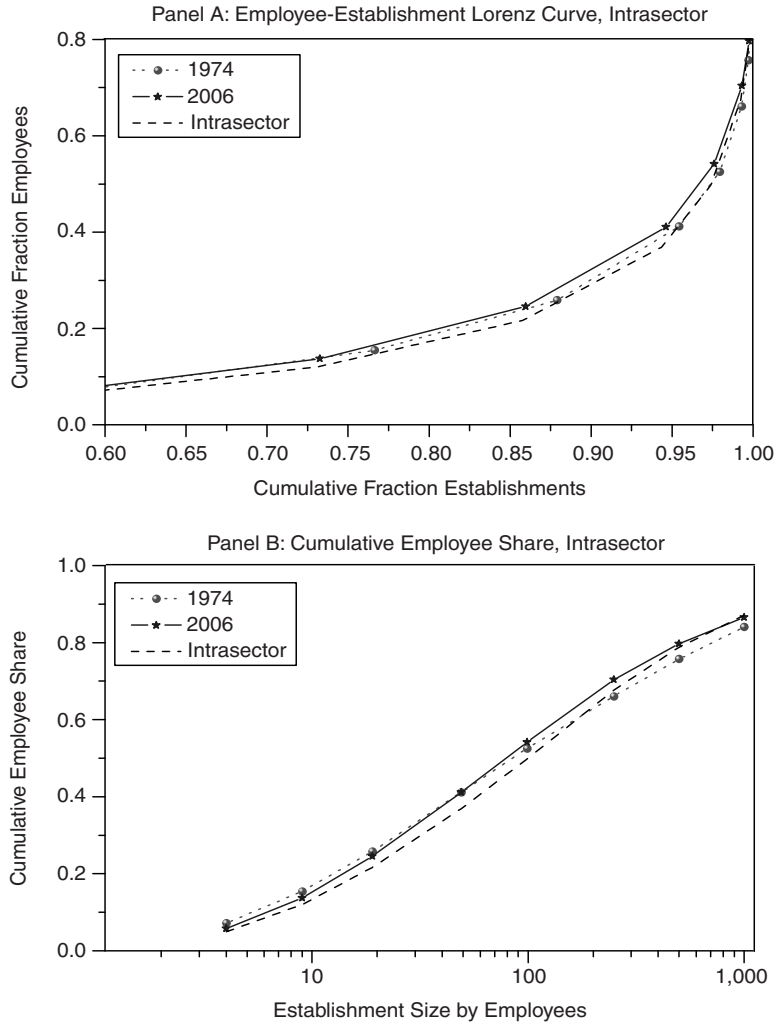
Table 2 presents the decomposition of trends in intra- and intersectoral changes. It shows how each statistic can be constructed as a weighted average of sectoral values. It also illustrates the computation of the counterfactual statistics used for the decomposition following the logic of equations (10) and (11). Considering only intrasector changes, the mean size of establishments would have increased 5 percent. Only the establishment mean of the manufacturing sector fell during this period, and its weight is relatively small. Keeping intrasector changes constant, the mean size would have dropped 12 percent. This is clearly because services, a sector with relatively small establishments in 1974, nearly doubled its share during this period.

Coworker mean results are substantially different. The main reason is that when labor shares are used instead of establishment shares, manufacturing is far more important than services. Consequently, when only intrasector changes are permitted, the drop in the coworker mean of manufacturing dominates the rise in services, and the coworker mean drops by 20 percent. Similarly, considering only intersector changes, the coworker mean size would have dropped 31 percent.[13] Finally, Table 2 presents the decomposition of the coefficient of variation of the establishment size distribution. The drop at the aggregate level is 7 percent. The decomposition shows that this drop is mainly due to intrasector changes. Keeping the share constant at 1974 levels, the drop would have been -14 percent; if one allows only changes in the share a fall of -2 percent is observed.

Figures 11 and 12 further resolve changes in the concentration of labor across establishments. Notice that these figures describe the distribution of workers across establishments, while the coefficient of variation mentioned earlier describes the distribution of establishments across establishment sizes. The results of this decomposition are different than those of the decomposition of the coefficient of variation. Allowing only intrasector changes, there would be a less equal distribution of labor across establishments in 2006 (see Figure 11). In contrast, intersector changes imply a greater shift toward a more even distribution than the one observed during this period.

[13] It is surprising in this case that with inter- or intrasector changes alone the coworker mean would have decreased more than when both changes occurred. This happens because the coworker mean size of services is higher than that of manufacturing in 2006, while the reverse is true in 1974. Thus, when the shares are allowed to change (not just the sectoral means), the aggregate coworker mean size increases.

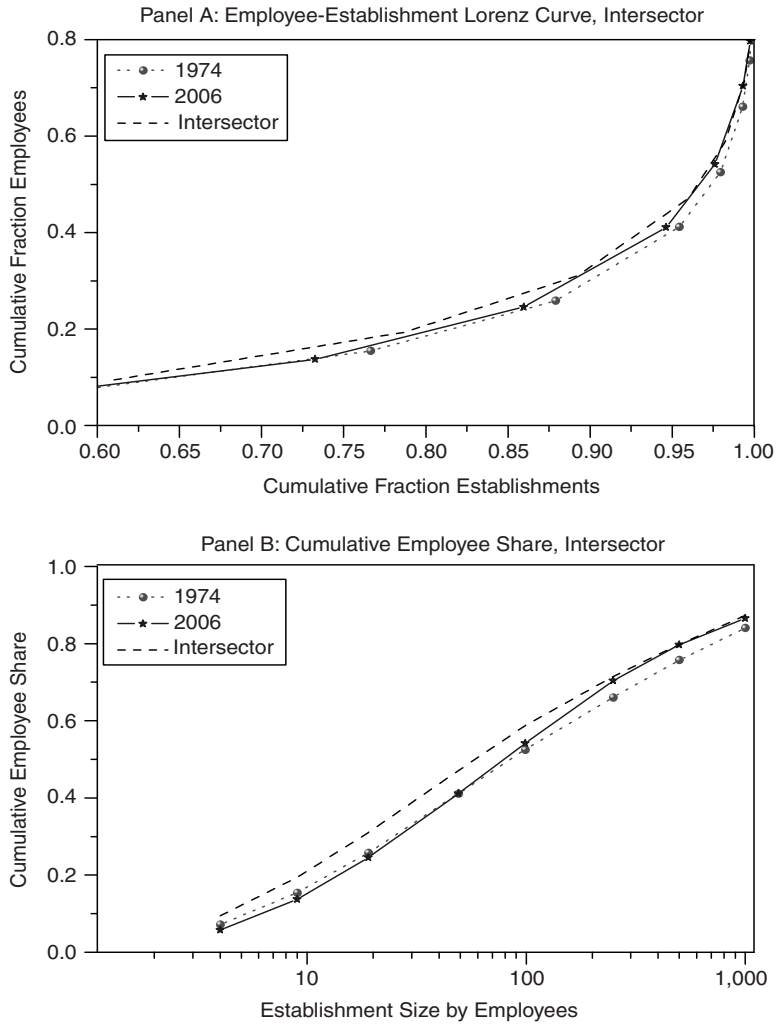
Figure 11 Intrasectorial Changes in the Establishment-Size Distribution, 1974–2006



4. FIRMS VERSUS ESTABLISHMENTS

Although the establishment is usually used as the production unit in models with heterogeneity in productivity, it is conceivable that the firm might also serve in that role. Because production units in these models vary in productivity or in their managers' ability, one could argue that they resemble establishments. However, since financial decisions are also made at the

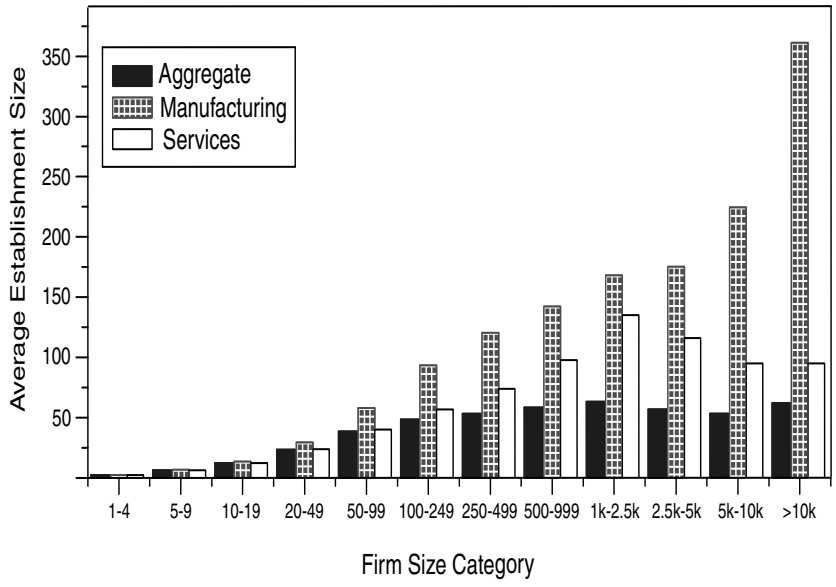
Figure 12 Intersectorial Changes in the Establishment-Size Distribution, 1974–2006



production unit level, it might also be argued that firm data is more appropriate. If it could be shown that the distribution of labor across firms tracks labor patterns across establishments, however, this distinction might be irrelevant.

It might be expected that small firms and small establishments, and large firms and large establishments, will see their labor distributions move together. Trivially, all small firms are composed entirely of small establishments, and all large establishments are constituent parts of large firms. If large firms contain

Figure 13 Establishment Size by Firm Size Group, 1991

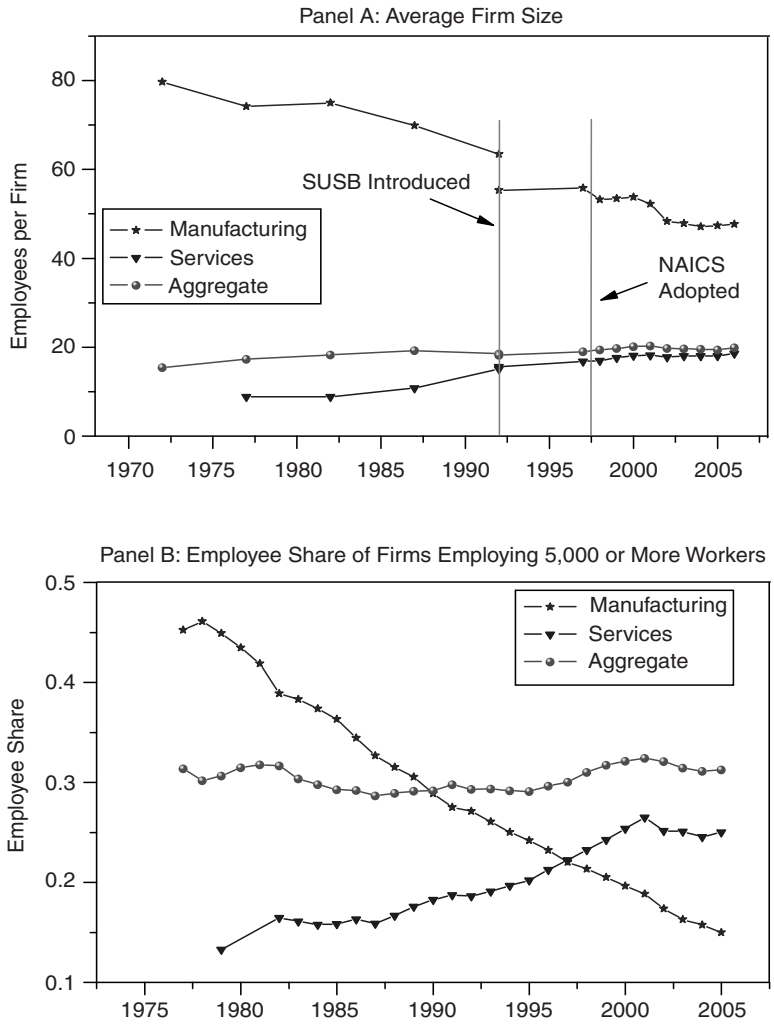


few small establishments, then the employment share of small establishments will correlate strongly with the employment share of small firms; the same will be true of large establishments and large firms. However, one may imagine a world in which large firms are mostly composed of many small establishments, and in this case movements in the distribution of labor across establishments might not be clearly reflected in movements of workers among firms. Consequently, it might be expected that co-movement in labor across establishments and across firms tends to be greater when large firms are composed of larger establishments.

Firm Data Sources

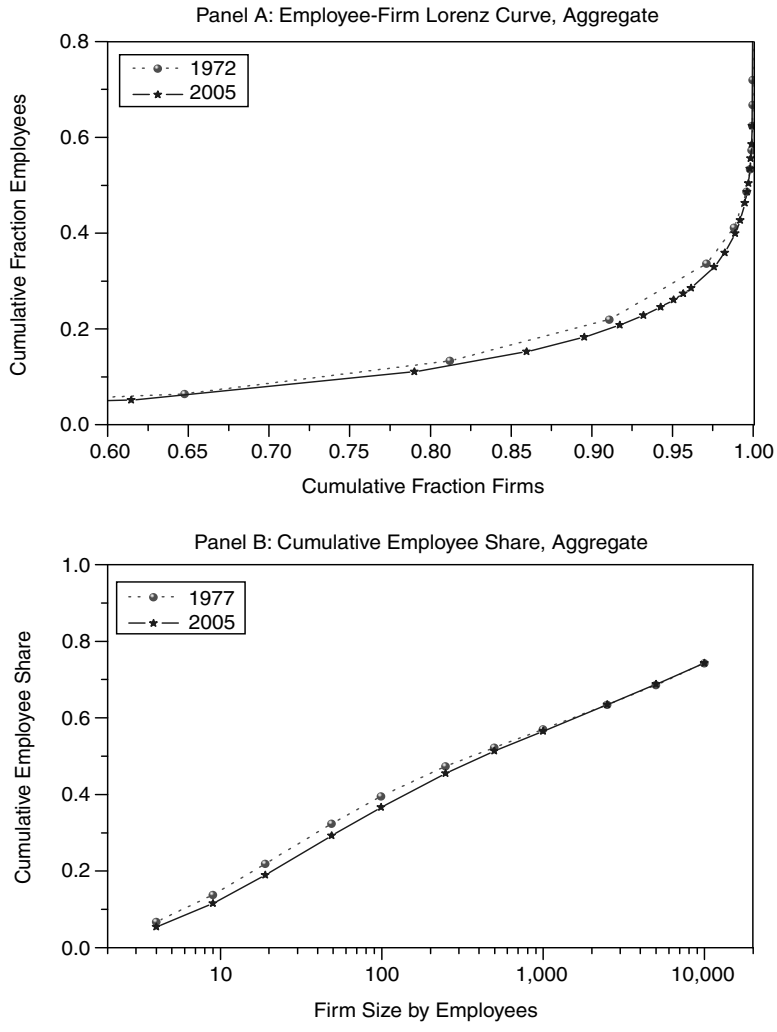
Firm data were obtained from three Census Bureau series: Enterprise Statistics, Statistics of U.S. Businesses (SUSB), and Business Dynamics Statistics (BDS). All series contain tallies of establishments and employees by firm size; Enterprise Statistics and SUSB also contain a count of firms in each firm size group. Enterprise Statistics was published consistently every five years from 1967 to 1992; SUSB was published in 1992 and annually after 1997. BDS was constructed retrospectively from several internal census databases and is available annually from 1977.

Figure 14 Firm Size and the Share of Large Firms in Total Employment



Whenever possible, BDS data are utilized. The publication is consistent in scope and methodology over the entire period of study. SUSB and especially Enterprise Statistics suffer from shifting definitions and sector coverage. These deviations, and the methods used in this article to mitigate their effects, are discussed in the Appendix.

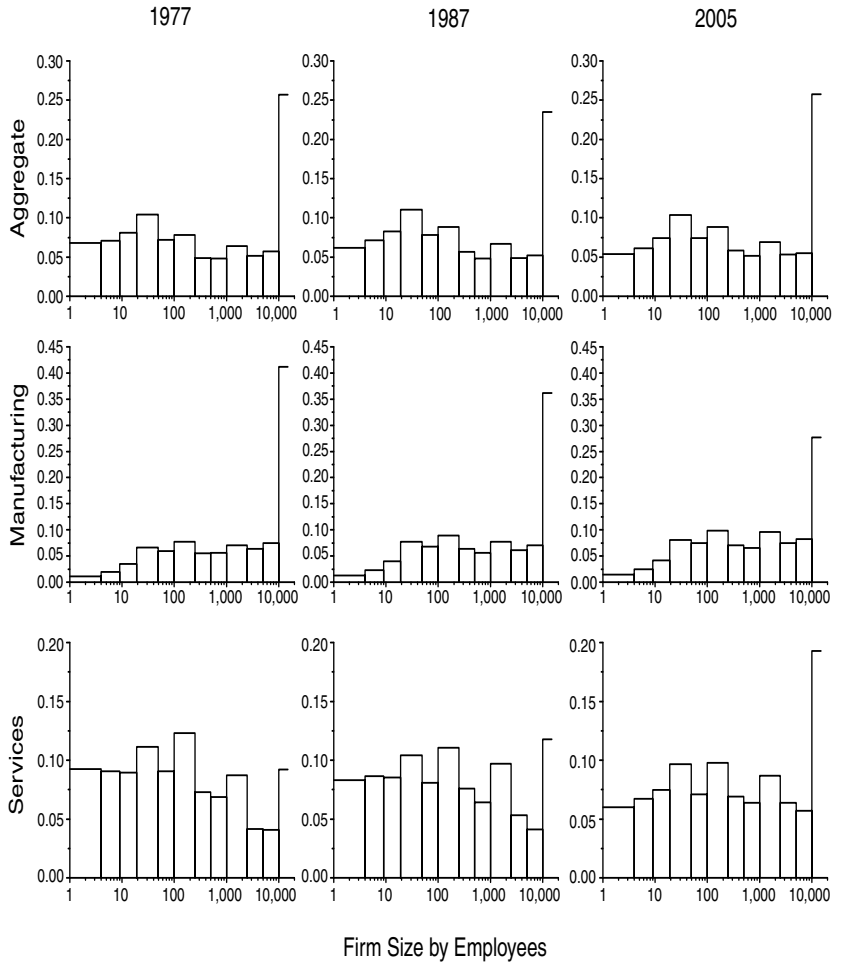
Figure 15 Firm-Size Distribution; Aggregate Economy, 1977–2006



Comparison Results

Figure 13 shows the average size of establishments for firms in 12 size categories in 1991; the data in this figure are typical for the sectors depicted and for the years 1979–2005. These data were obtained from BDS. Large firms, unlike small firms, do seem to be composed of larger establishments, and this is even more true in the manufacturing sector than in the rest of the economy. Movements in labor distribution should be similar across establishments

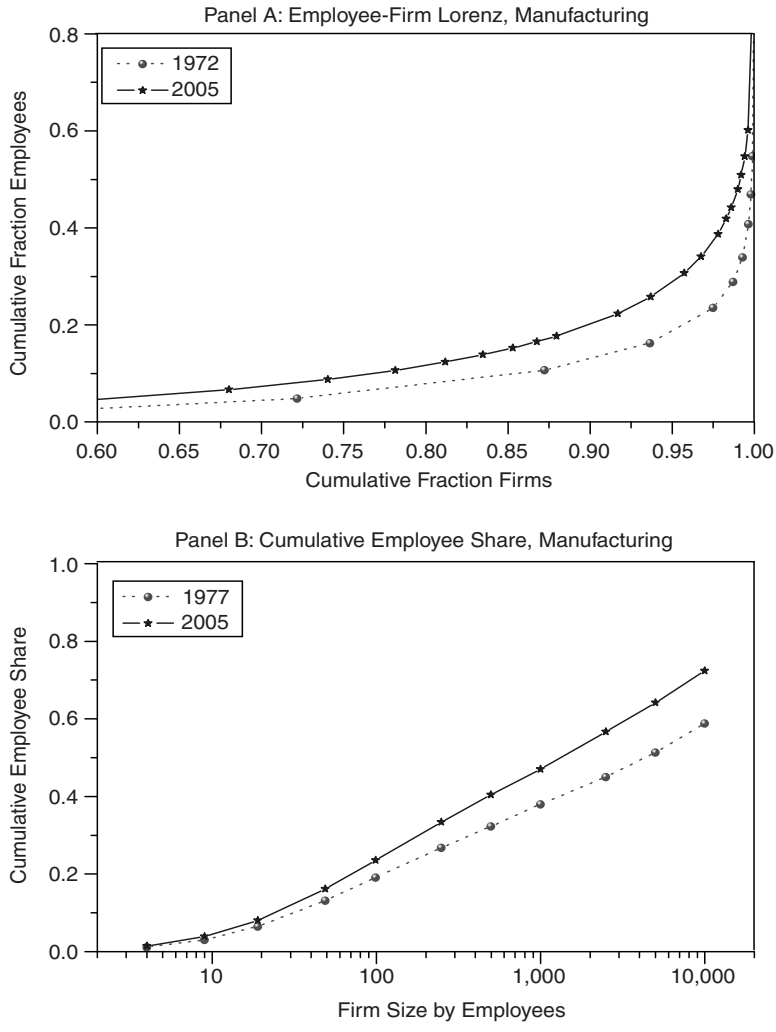
Figure 16 Histograms for the Distribution of Labor Across Firms



and firms, then, especially within the manufacturing sector. Indeed, evidence presented below generally confirms firm-establishment labor co-movement in these sectors, and to a degree in the aggregate economy, at least in the period under examination.

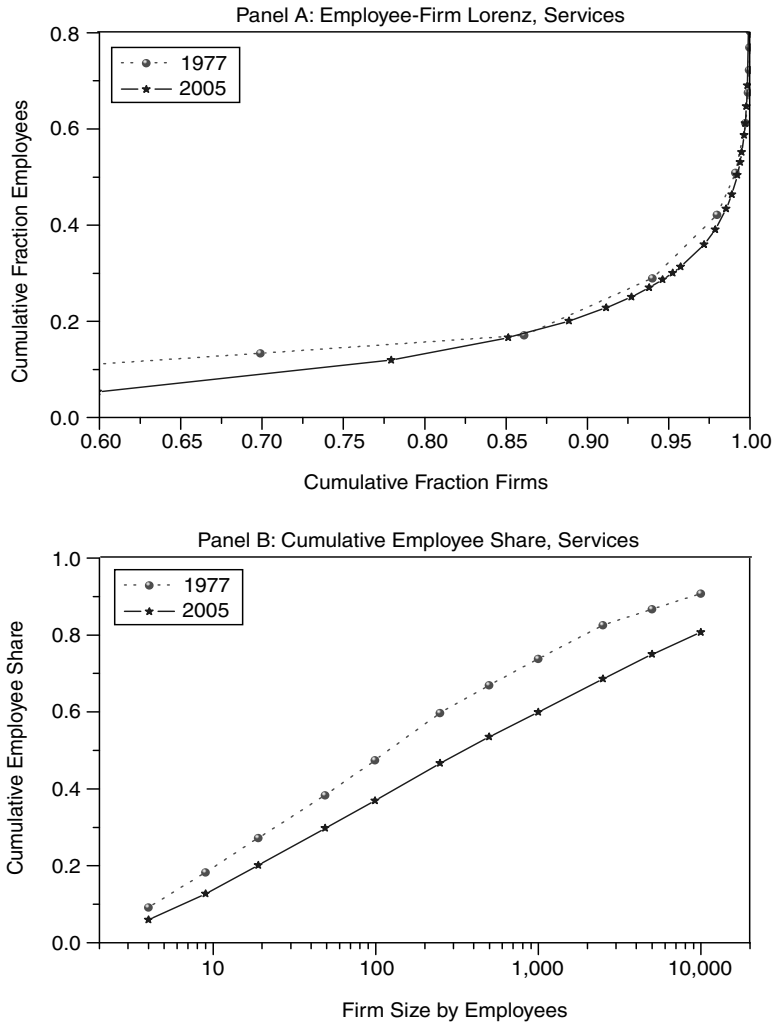
Figures 14 through 18 display firm data analogous to the establishment data. Data used in the creation of Lorenz curves (Panel A in Figures 15, 17, and 18) and mean firm size series (Figure 14, Panel A) were obtained through Enterprise Statistics and SUSB. Other firm figures (Panel B in Figures 14, 15, 17, and 18, as well as all of Figures 13 and 16) were derived from the BDS series.

Figure 17 Firm-Size Distribution; Manufacturing Sector, 1974–2006



It is clear that labor distribution movements across establishments track those in firms. Both the aggregate and the sectoral mean size series display the same patterns between the early 1970s and mid-2000s that are seen at the establishment level. Intrasector changes in the distribution of employment by firm size resemble those in establishment data: labor in the manufacturing sector became less concentrated (more clearly for firms than establishments), while service sector labor grew slightly more concentrated. Perhaps the only qualitative departure from establishment trends is a decrease in the

Figure 18 Firm-Size Distribution; Service Sector, 1977–2006



evenness of the aggregate labor distribution across firms that occurred between 1972 and 2005.

5. CONCLUSIONS

This article collects and analyzes publicly available data from the 1970s onward to obtain a set of statistics that can be used to calibrate and evaluate models with establishment heterogeneity. Recently, these models have

become widely used in economics to explain phenomena as important as economic development.

At the aggregate level, there is a minor shift of labor to mid-size establishments and away from the smallest and largest establishments. This change is partially explained by intrasector changes. The largest manufacturing establishments have consistently lost employee share since 1974, and manufacturing establishments smaller than 500 employees have uniformly seen their employee share increase. Trends in the distribution of labor across service sector establishments are complicated by inconsistencies in the definition of the service sector, but service establishments seem to have become larger since 1974, and the largest service establishments have grown at a disproportionately fast rate. Thus, the distribution of labor across service establishments has become less even, with most change occurring before 1997. Changes in the aggregate distributions of establishments and labor across establishments are also the result of changes in the share of sectors. Between 1974 and 2006 the worker share of manufacturing, a sector with large establishments and concentrated labor, decreased as the employment share of the services sector, characterized by smaller establishments, increased. In combination with movements in intrasector distributions, this trend explains observed changes in the aggregate distributions of establishments and labor across establishments.

Labor movements across firms should, hypothetically, resemble the movement of labor across establishments. This will be true to a greater degree when large firms contain fewer small establishments. This hypothesis is not contradicted by the data presented in this article.

APPENDIX

Data Sources

Enterprise Statistics

The Enterprise Statistics (ES) data set was first published in 1954; later publications came in 1958, 1963, 1967, and every five years after 1967 until the series was discontinued after 1992. The primary virtue of ES for this article is the provision of tables detailing quantities of firms, establishments, and employment; these values are provided for firms in different employment size groups similar to establishment size groups in CBP. These size groups are available for the aggregate economy as well as for sectors that generally replicate SIC definitions.

Unfortunately, ES's coverage and content changes significantly from publication to publication. The number of SIC sectors covered varies wildly;

using sector-level data we were able to homogenize the aggregate data, but the adjusted series lacks coverage of entire sectors (transportation and communication; finance and real estate; and most services). Moreover, the 1972 publication inflates its count of small firms by including certain non-employers; this can be corrected for the aggregate using a table found in that publication's appendix. The manufacturing sector from this year is still usable because there are no manufacturing firms in the small size group affected by the 1972 methodology, but the sector-level data for service firms must be set aside.

Adjustment of ES data to obtain a homogenous aggregate composition requires the subtraction of some sectors from each year's aggregate. This is a simple arithmetic task complicated in some cases by the lack of subsector data: The Census Bureau occasionally withholds employment information for certain firm size groups if its publication might result in the disclosure of private information. These missing values are estimated by multiplying the number of firms in the size group with the missing data by the mean number of employees per firm for the size group at the aggregate level. An example adjustment is displayed in Table 3. There, the original employee count for each aggregate firm size group was reduced by the deduction of employees in public warehousing, travel agencies, and dental laboratories—three small sectors not present in the ES aggregates in all years. Values in bold were missing from the original publication and estimated using the procedure previously described. Similar exercises were also carried for firm and establishment series and in all ES years.

The composition of the services sector also varied from publication to publication. Unfortunately, homogenization was not a feasible solution: very few firms would remain in an intertemporally consistent services sector. Consequently, the service sector is presented for each ES year unaltered with the caveat that it is inconsistent.

Statistics of United States Businesses

Statistics of U.S. Businesses (SUSB) replaced ES in 1992; it was published in 1992, and annually from 1997 onward. Although SUSB provides data similar to those found in ES, there are several important differences. First, SUSB covers many sectors not covered by ES. This leaves aggregate data somewhat incomparable across the two publication series, especially after this article's sectoral homogenization of aggregate ES data. Second, SUSB uses enterprise size groups rather than firm size groups. In ES these terms were interchangeable and each enterprise was assigned a single industry code; in SUSB an enterprise is composed of many firms, each of which represents the enterprise's production in a given industry. With this convention, it is possible to find a 5,000–9,999 employee size group containing three firms employing 2,000 workers between them. This data is not well-suited for the creation of Lorenz curves because it does not permit the sorting of firms by size.

Table 3 Adjustment to ES Sectoral Composition; Example

Firm Size Group	Original Total Total Employees	(Subtracted) Public Warehousing (42A)	(Subtracted) Travel Agencies (47)	(Subtracted) Dental Laboratories (80)	Final Figure Adjusted Total Employees
0	0	0	0	0	0
1–4	2,938,355	5,235	8,898	6,355	2,917,867
5–9	3,209,609	8,471	10,447	5,107	3,185,584
10–19	3,945,190	14,670	7,869	5,284	3,917,367
20–49	5,372,937	27,508	5,997	6,072	5,333,360
50–99	3,446,571	13,739	3,100	1,670	3,428,062
100–249	3,459,628	14,281	1,967	1,526	3,441,854
250–499	2,126,488	**5,833**	2,100	**686**	2,117,869
500–999	1,837,286	**688**	**688**	0	1,835,910
1,000–2,499	2,330,673	**4,618**	**3,079**	**1,539**	2,321,437
2,500–4,999	1,981,793	0	0	0	1,981,793
5,000–9,999	2,376,041	0	0	0	2,376,041
= 10,000	12,786,233	0	0	0	12,786,233
Total	45,810,804	94,464	44,888	27,744	45,643,708
Column Error	0	−579	743	−496	331

Values in bold were missing from the original publication and are estimated using the procedure described in the text of this article.

Table 4 Services Sector Assembled from NAICS

NAICS Number	NAICS Service Sector Component
54	Professional, scientific, and technical services
56	Administrative and support and waste management and remediation services
61	Educational services
62	Health care and social assistance
71	Arts, entertainment, and recreation
72	Accommodation and food services
81	Other services (except public administration)

Moreover, it prevents any adjustment of the SUSB aggregate by the subtraction of sector data, because too many firms would be dropped. For example, if the construction and mining sectors are subtracted from the aggregate, and a single enterprise has constituent firms in each sector, then two firms will be removed from the aggregate despite the fact that the enterprise is represented in the aggregate by a single firm. Consequently, sectoral and aggregate data are only marginally comparable between the two series.

The utility of SUSB is further reduced by the switch to the NAICS classification system from the SIC system after 1997; it is difficult to compare sectors between systems, and, as with CBP, it was necessary to construct a composite service sector from several NAICS subsectors (see Table 4). Because of the SUSB definition of a firm, the number of service firms in large size groups is probably overstated in NAICS.

Business Dynamics Statistics

BDS is consistent in methodology and coverage; derived from a number of internal USCB databases, it has annual data on employment for firm size groups reaching back to 1977. For the purposes of this article, BDS has one major shortcoming: For each firm size group, only data on establishments and employment are provided. When firm quantities are required for a calculation, ES and SUSB are used.

Because the series was assembled from microdata retrospectively, BDS industry classifications are internally comparable for all years. These classifications are based on the SIC system, and so the comparability of BDS sector data with CBP and SUSB sector series from 1998 on is somewhat compromised.

Computing Establishment and Coworker Means and Probabilities

We compute the expected establishment mean for a size group by dividing the total number of workers in a size group ($workers_i$) by the total number of establishments in the size group ($establishments_i$):

$$E(esize \mid egroup = i) = \frac{workers_i}{establishments_i}. \tag{12}$$

Obtaining the expected coworker mean for a size group is more involved and the next subsection is devoted to this effort. Meanwhile, the probabilities $P(egroup = i)$ and $P(wgroup = i)$ are obtained by dividing the establishments or workers (respectively) in i by the total number of establishments or workers over all size groups j:

$$P(egroup = i) = \frac{establishments_i}{\sum_j establishments_j}, \text{ and} \tag{13}$$

$$P(wgroup = i) = \frac{workers_i}{\sum_j workers_j}. \tag{14}$$

Probabilities $P(egroup \leq i)$ and $P(wgroup \leq i)$ are calculated in a similar manner by summing the probabilities for each size group j less than or equal to i:

$$P(egroup = i) = \frac{\sum_1^i establishments_j}{\sum_j establishments_j}, \text{ and} \tag{15}$$

$$P(wgroup = i) = \frac{\sum_1^i workers_j}{\sum_j workers_j}. \tag{16}$$

Computing the Size-Group Coworker Mean

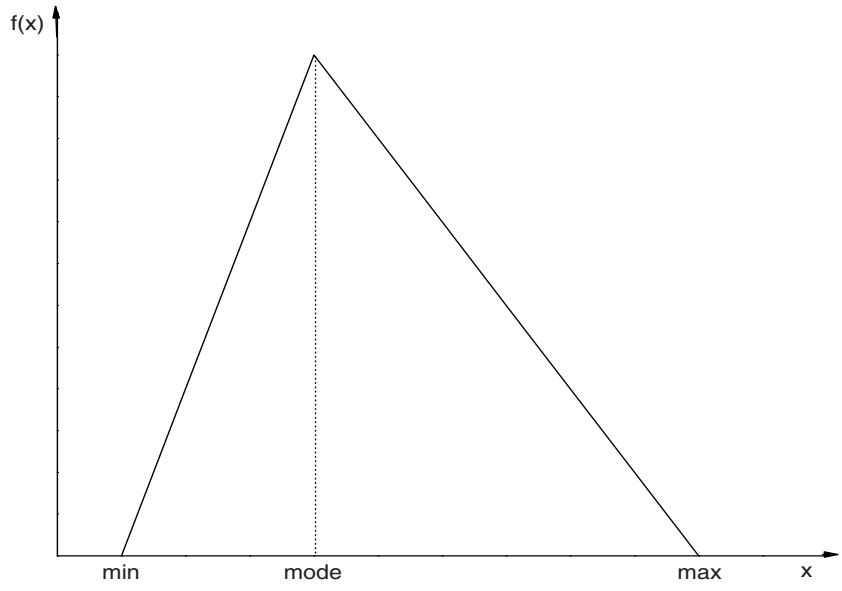
For each size group i, the available information is

- the minimum and maximum size in the group, min_i and max_i, respectively;
- the total number of workers, $workers_i$; and
- the total number of establishments, $establishments_i$.

With this information it is simple to compute the mean size of the group,

$$E(esize \mid egroup = i) = \frac{workers_i}{establishments_i}. \tag{17}$$

Figure 19 Triangular Distribution; Example



Unfortunately, it is not possible to compute the coworker mean of this group. Davis and Haltiwanger (1989) show that the coworker mean can also be written as

$$E(wsize \mid wgroup = i) = E(esize \mid egroup = i) + \frac{V(esize \mid egroup = i)}{E(esize \mid egroup = i)}, \tag{18}$$

where $V(esize \mid egroup = i)$ is the variance of the establishment size for the size group i. Equation (18) indicates that once $E(esize \mid egroup = i)$ is known, only an estimate of $V(esize \mid egroup = i)$ is needed to obtain an estimate of $E(wsize \mid wgroup = i)$. With a distributional assumption for the distribution of establishments inside each size group, this statistic can be recovered. A useful assumption is that this distribution is triangular. This distribution has three parameters: the lower bound, min; the upper bound, max; and the mode, $mode$. The probability density function increases linearly from min to $mode$ and decreases linearly from $mode$ to max (see Figure 19 for an example). With this assumption, the mean size can be written as

$$E(esize \mid egroup = i) = \frac{min_i + max_i + mode_i}{3}. \tag{19}$$

Since $E(esize \mid egroup = i)$, min_i, and max_i are available, one can use the equation above to solve for $mode_i$. Then, it is simple to compute the variance using the formula for the triangular distribution,

$$V(esize \mid egroup = i) =$$

$$\frac{min_i^2 + max_i^2 + mode_i^2 - min_i * max_i - min_i * mode_i - max_i * mode_i}{18}. \tag{20}$$

Finally, equation (18) can be used to compute the coworker mean of size group i.

REFERENCES

Alfaro, Laura, Andrew Charlton, and Fabio Kanczuk. 2008. "Plant-Size Distribution and Cross-Country Income Differences." Working Paper 14060. Cambridge, Mass.: National Bureau of Economic Research.

Amaral, Pedro S., and Erwan Quintin. 2007. "Limited Enforcement, Financial Intermediation, and Economic Development: A Quantitative Assessment." Manuscript, Federal Reserve Bank of Dallas.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293 (September): 1,818–20.

Banerjee, Abhijit V., and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven Durlauf. Amsterdam: Elsevier, 473–552.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2008. "Cross-Country Differences in Productivity: The Role of Allocative Efficiency." Manuscript, University of Maryland.

Buera, Francisco J., and Joseph P. Kaboski. 2008. "Scale and the Origins of Structural Change." Federal Reserve Bank of Chicago Working Paper 2008-06.

Buera, Francisco J., and Yongseok Shin. 2008. "Financial Frictions and the Persistence of History: A Quantitative Evaluation." Mimeo, Northwestern University.

Caselli, Francesco, and Nicola Gennaioli. 2003. "Dynastic Management." Working Paper 9442. Cambridge, Mass.: National Bureau of Economic Research.

Castro, Rui, Gian Luca Clementi, and Glenn McDonald. 2009. "Legal Institutions, Sectoral Heterogeneity, and Economic Development." *Review of Economic Studies* 76 (April): 529–61.

Davis, Steven J., and John Haltiwanger. 1989. "The Distribution of Employees by Establishment Size: Patterns of Change in the United States, 1962–1985." Manuscript, University of Chicago and University of Maryland.

Davis, Steven J., and John Haltiwanger. 1990. "Size Distribution Statistics from County Business Patterns Data." Manuscript, University of Chicago.

Davis, Steven J., John Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, Mass.: The MIT Press.

Gibrat, R. 1931. *Les Inégalités Économiques; Applications: Aux Inégalités des Richesses, à la Concentration des Entreprises, Aux Populations des Villes, Aux Statistiques des Familles, etc., d'une Loi Nouvelle, La Loi de l'Effet Proportionnel*. Paris: Librarie du Recueil Sirey.

Greenwood, Jeremy, Juan M. Sánchez, and Cheng Wang. 2008. "Financing Development: The Role of Information Costs." Federal Reserve Bank of Richmond Working Paper 08-08.

Guner, Nezih, Gustavo Ventura, and Xu Yi. 2008. "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics* 11 (October): 721–44.

Hopenhayn, Hugo, and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101 (October): 915–38.

Hsieh, Chang-Tai, and Peter J. Klenow. 2007. "Misallocation and Manufacturing TFP in China and India." Working Paper 13290. Cambridge, Mass.: National Bureau of Economic Research (August).

Lucas, Jr., Robert E. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (Autumn): 508–23.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics* 11: 707–20.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. 2009. *Integrated Public Use Microdata Series: Version 4.0*. Minneapolis: Minnesota Population Center. http://usa.ipums.org/usa/.

Sutton, John. 1997. "Gibrat's Legacy." *Journal of Economic Literature* 35 (March): 40–59.