measures. Thus, small employers also destroy jobs at a much higher rate than large employers.

How does net job creation vary by employer size? On this score, the empirical evidence produces no strong pattern. Net job creation rates by firm size exhibit a ∩ shape: manufacturing firms with 100–499 employees show mild net contraction rates during the 1973–88 period, whereas smaller and larger firms show sharper contraction rates. Neither plant-size measure evinces any strong relationship to net job creation rates, although the net contraction rate is substantially smaller for single-unit than multi-unit firms. In a nutshell, net job creation in the U.S. manufacturing sector exhibits no strong or simple relationship to employer size.

How can we reconcile this empirical result with the widely held belief that small businesses account for a disproportionate fraction of new jobs? One might think that the answer lies in our focus on the manufacturing sector. Perhaps in the nonmanufacturing sectors of the economy, smaller firms exhibit much higher net job creation rates than do larger firms. It is conceivable that an analysis of nonmanufacturing sectors along the lines of table 4.1 would uphold the conventional wisdom about the role of small business in job creation, but it is not this type of analysis that underlies the conventional wisdom. Rather, the conventional wisdom rests on fallacious and misleading interpretations of the data, as we explain in the next three sections.

4.4 The Size Distribution Fallacy

Many claims about the job-creating prowess of small business appear to be based upon changes over time in the size distribution of employment. We review the calculation typically performed on the size distribution data and explain why the usual interpretation of this calculation leads to fallacious inferences about job creation.

The SBA defines small businesses as firms with fewer than 500 employees, although the precise cutoff is not important to the point at hand. Given a particular cutoff, let $TOTAL_t$ and $SMALL_t$ stand for total employment and small business employment, respectively, in year t. In terms of these symbols, one can calculate the small business "contribution" to 1990 job creation as the ratio

$$\frac{SMALL_{1990} - SMALL_{1989}}{TOTAL_{1990} - TOTAL_{1989}}.$$

Box 4.1
Illustration of the size distribution fallacy

	Firm 1	Firm 2	Firm 3	Small firms	Big firms	All firms
Year 1 employment	300	550	650	300	1,200	1,500
Year 2 employment	50	340	1,210	390	1,210	1,600
Net change	−250	−210	560	90	10	100

Small-firm share of net job creation = (390 − 300)/(1,600 − 1,500) = .9

This illustration uses data in the three rightmost columns on the size distribution of employment to calculate job creation shares. These changes in the distribution of employment by firm size ignore the fact that firms can migrate between size categories, as shown in the three leftmost columns, resulting in a false inference about the share of job creation accounted for by small firms.

In words, the small business contribution to 1990 job creation is equated to the ratio of net employment change among small firms to total net employment change.[8]

The fallacy arises because firms can migrate between size categories from one year to the next. The example in box 4.1 illustrates this point. The example considers three firms, one of which (firm 1) satisfies the SBA definition of a small business in year 1. The largest firm (firm 3) grows dramatically in year 2, whereas the two smaller ones shrink. As it shrinks, firm 2 migrates from the large-business to the small-business category. On net, total employment increases by 100.

If one executes the typical calculation on data in the example, small business appears to contribute 90 percent of net job growth. But, as the construction of the example makes clear, this interpretation is fallacious. In the example, firm-level net job growth actually increases with firm size, an observation that can be made only by following individual employers over time, as in the calculations that underlie the net and gross job flow figures in table 4.1.

How important is such migration across firm size categories in reality? The large magnitude of gross job flows—and the concentration of job flows in plants that undergo big employment changes—indicates that migration

8. Zayas (1978) uses data on changes in the size distribution of employment to calculate growth rates by size of business. This calculation is also subject to the size distribution fallacy identified below.

across categories is frequent and important.[9] Especially during periods of slow employment growth, firm migration from large to small is likely to occur quite often. This pattern creates the appearance of a booming small-firm sector.

In summary, many claims about the job-creating prowess of small businesses derive from a fallacious interpretation of data on the size distribution of employment. Size distribution data cannot tell us whether small businesses systematically grew faster than large businesses.

4.5 Netting Out Reality

Sophisticated proponents of the view that small businesses create a disproportionate fraction of new jobs recognize the fallacy described above.[10] Circumventing the fallacy requires longitudinal data on individual establishments or firms—that is, data that track individual employers over time. The most widely cited studies of job creation behavior rely upon such data, but they often present results in a way that can mislead the statistically naive.[11]

To understand the potential for confusion, consider the example in box 4.2. It depicts a situation with moderate net job growth in the midst of much larger gross job flows. We know from table 4.1 that this situation typifies the experience of the U.S. manufacturing sector. It also typifies the experience in other sectors of the U.S. economy and in other industrialized nations.[12]

In the example, 100 percent of the net job increase between years 1 and 2 is accounted for by firm 1, which is classified as small on the basis of its employment in year 1. Thus, one might conclude that "small firms created virtually all new jobs" between years 1 and 2. Closer analysis reveals, however, that such a conclusion grossly mischaracterizes the distribution of newly created jobs by size of firm. In fact, in this example large firms create 80 percent of the new jobs in year 2.

Public discourse about job creation rarely distinguishes between the small business share of gross job creation (20 percent in the example) and its "share"

9. Table 4.1 shows that gross job creation and destruction flows are large relative to net employment changes. Chapter 2 contains additional evidence on this point and on the concentration of gross job flows at plants that undergo big employment changes.

10. Small Business Administration (1983, p. 62) clearly explains the fallacy. See also Birch and MacCracken (1983).

11. The most widely cited studies of the small-business role in creating jobs are Small Business Administration (various years) and Birch (1979, 1987).

12. See chapters 2 and 3 for a review of the evidence.

Box 4.2
Illustration of a confusion between net and gross job creation

	Firm 1	Firm 2	Firm 3	Small firms	Big firms	All firms
Year 1 employment	300	600	600	300	1,200	1,500
Year 2 employment	350	400	800	350	1,200	1,550
Net change	50	−200	200	50	0	50

Small-firm share of net job creation = $50/50 = 1$
Small-firm share of gross job creation = $50/(50 + 200) = 0.2$

This illustration calculates job creation shares from longitudinal data on individual firms. The calculation makes use of longitudinal data to calculate net firm-level employment changes. Those employment changes are aggregated over firms within a size class and then expressed as a fraction of the aggregate net change. Following the common practice of prominent analysts and government agencies like the U.S. Small Business Administration, continuing firms are assigned to a size category by using base-year employment. The last two lines show how the small-firm share of net job creation misrepresents the actual distribution of newly created jobs by size of firm.

of net job creation. Consequently, claims about the job creation role of small business often conjure up the image of an economy in which large firms inexorably shrink and small firms struggle valiantly to replenish the stock of jobs. This image deviates sharply from the facts set out in table 4.1 and in table 4.3, which show that both large and small employers create large numbers of new jobs.

To appreciate fully the misleading character of statements about the small business "share" of net job creation, consider a particular historical episode. Between March 1973 and March 1974, manufacturing employment as reported in the LRD increased on net by about 16,000 jobs. Over this same period, manufacturing plants with fewer than 100 employees as of March 1973 experienced a net increase of about 160,000 jobs. Thus the net increase for small plants was ten times the overall net increase. If we summarized these data in the usual phraseology of public discourse, we would say that "small employers created 1,000 percent of the new manufacturing jobs in 1974." Proponents of the small-business job creation view would likely eschew the usual phraseology in this case, because it highlights the absurdity of the underlying calculation.

Continuing with the historical episode, manufacturing plants of more than 500 employees created about 1.3 million gross new jobs between 1973 and 1974. Since net job growth was only 16,000 during this period, we could easily identify a set of large manufacturing plants that accounted for 50 percent, 100 percent, 200 percent, or 1,000 percent of net job growth. We could do so by choosing a set of large plants situated in states with robust employment growth or rapidly expanding industries. We could even identify several distinct sets of large plants, each of which accounted for, say, 100 percent of net job growth. It is unlikely that useful economic policy prescriptions would follow from these characterizations of the data. Yet it is precisely this type of data characterization and argument that underlies claims that small businesses create most jobs and—therefore—ought to receive favorable tax and regulatory treatment.

In summary, longitudinal studies that focus on the "share" of net job growth accounted for by small businesses grossly misrepresent the actual distribution of newly created jobs by size of employer. A more meaningful way to represent this distribution is to focus on the small-employer share of gross job creation.[13]

4.6 The Regression Fallacy

Most longitudinal studies of the relationship between employer size and job creation suffer from another statistical pitfall known as the regression fallacy or regression-to-the-mean bias.[14] The potential for bias arises whenever employers experience transitory fluctuations in size, or whenever measurement error introduces transitory fluctuations in observed size. Both phenomena are important features of longitudinal data on employers.

The simple example in box 4.3 illustrates the regression fallacy. It calculates growth rates for individual firms and by size of firm for years 2 and 3. Following widespread practice, firms are assigned to size classes using base-year employment.[15] The base year is the initial year of the time interval over which a particular growth rate is calculated.

Boldface entries in the illustration represent average employment growth rates by size class in years 2 and 3. These entries convey the impression that small firms outperform large ones in both years. Yet, closer inspection reveals that each firm is the same size in year 3 as in year 1. Evidently, the seemingly appropriate calculations underlying the boldface entries provide a misleading characterization of the size-growth rate relationship. This misleading characterization is an example of the regression fallacy.

The fallacy arises because, each year, we reclassify firms into size classes using base-year employment. The interaction between this reclassification and transitory firm-level employment movements lies at the heart of the regression fallacy. On average, firms classified as large in the base year are more likely to have experienced a recent transitory increase in employment. Since transitory movements reverse themselves, firms that are large in the base year are relatively likely to contract. Likewise, firms classified as small in the base year are

13. For the record, we should note that not every statistical tabulation performed on longitudinal data by the SBA examines the small-employer share of net job creation. For example, table 13 in Small Business Administration (1988) reports gross job creation by firm size. Nonetheless, the surrounding text reverts to the misleading "net" calculation when characterizing the small-business role in job creation.

14. Friedman (1992) suggests that the regression fallacy "is the most common fallacy in the statistical analysis of economic data." Leonard (1986) explains how regression-to-the-mean bias can distort the estimated relationship between employer size and growth rates. Friedman (1992) and Quah (1992) focus on the regression fallacy in the recent literature that investigates whether per capita income levels are converging across countries.

15. This classification practice is used, for instance, in the annual SBA reports to the president and in Birch (1979, 1987).

Box 4.3
Illustration of the regression fallacy

	Firm 1	Firm 2	Firm 3	Small firms	Big firms	All firms
Year 1 employment	450	550	600	450	1,150	1,600
Year 2 employment	550	450	600	450	1,150	1,600
Year 3 employment	450	550	600	450	1,150	1,600
Year 2 growth rate	.22	−.18	0	.22	−.09	0
Year 3 growth rate	−.18	.22	0	.22	−.09	0

This illustration calculates net job creation rates for individual firms and by size class of firms. Following the common practice of prominent analysts and government agencies like the U.S. Small Business Administration, continuing firms are assigned to a size category using base-year employment. Year 1 (year 2) is the base year when calculating year 2 (year 3) growth rates. Although each firm employs the same number of workers in year 1 as in year 3, the net growth rate for small firms—as calculated—exceeds the net growth rate for big firms in years 2 and 3. This apparent puzzle reflects a bias in the estimated size-growth relationship induced by temporary changes in the level of employment at individual firms.

more likely to have experienced a recent transitory decrease in employment. Hence, firms that are small in the base year are relatively likely to expand. As in our illustration, this regression phenomenon (i.e., regression to the firm's long-run size) creates the illusion that small firms systematically outperform large ones.

The magnitude of the bias associated with the regression fallacy depends on several factors: the extent of measurement error in the data, the importance of transitory employment movements for individual employers, the size distribution of employment, and the precise size-class boundaries chosen by the analyst. As a consequence, we cannot precisely quantify the extent of regression-to-the-mean bias in previous studies without direct access to their longitudinal data. We can, however, replicate their procedure for measuring employer size in the LRD and determine the resulting relationship between size and net job growth. We can then compare this size-growth relationship with those that emerge under alternative size measures.

Table 4.2 carries out this comparison, using LRD data for the period 1973–88. Following the standard (Birch/SBA) practice described above, the first panel classifies continuing plants and plant deaths by base-year size. New plants are classified according to size in the entry year. As we have explained, the entries in this panel are subject to the regression fallacy. To avoid the regression fallacy, we measure employer size by using average plant size or current plant size. Recall that current size equals the simple average of the plant's employment in the current and previous years, and average size equals a mean computed over all sample obervations on the plant.[16] Repeating portions of table 4.1, the bottom two panels of table 4.2 display the figures for average and current plant-size measures.

The results of the comparison are striking. In the first panel, the net job creation rate declines steeply over the first five size-class intervals and then flattens out over the remaining intervals. The second panel presents a sharp contrast. It indicates that the net job creation rate shows no systematic relationship to average plant size. The third panel actually shows a positive relationship between net job creation and current plant size. The gross job creation and destruction patterns also look much more favorable for small plants under the base-year size measure (first panel) than under either

16. To the extent that transitory employment fluctuations require more than one year to reverse themselves, our current-size measure is subject to a milder and more subtle version of the regression fallacy. However, random errors in measuring employment levels do not produce a regression fallacy under any of our plant or firm size measures.

Table 4.2
Net and Gross Job Flow Rates by Three Measures of Plant Size: Mean Annual Rates, 1973 to 1988

	Job Creation	Job Destruction	Net Growth	Employment Share
Birch/SBA Plant-Size Measure				
0 to 19 Employees	25.7	15.4	10.3	5.2
20 to 49	13.6	13.1	0.6	8.5
50 to 99	11.4	12.0	−0.7	10.4
100 to 249	9.5	11.1	−1.7	18.6
250 to 499	7.4	9.9	−2.5	16.0
500 to 999	6.3	9.0	−2.7	13.5
1,000 to 2,499	5.7	8.4	−2.6	12.3
2,500 to 4,999	5.4	7.9	−2.5	7.0
5,000 or more	4.7	7.1	−2.4	8.5
Average Plant-Size Measure				
0 to 19 Employees	15.9	17.2	−1.3	4.4
20 to 49	12.6	13.8	−1.1	8.2
50 to 99	11.7	12.6	−0.9	10.1
100 to 249	10.0	11.5	−1.4	18.5
250 to 499	8.5	9.8	−1.3	16.6
500 to 999	7.5	8.5	−1.0	13.8
1,000 to 2,499	6.6	8.2	−1.6	12.5
2,500 to 4,999	6.5	8.2	−1.7	7.2
5,000 or More	5.9	6.5	−0.6	8.8
Current Plant-Size Measure				
0 to 19 Employees	18.7	23.3	−4.5	5.2
20 to 49	13.2	15.3	−2.1	8.6
50 to 99	12.2	13.5	−1.3	10.5
100 to 249	9.6	10.7	−1.1	18.5
250 to 499	7.7	8.7	−1.0	16.0
500 to 999	7.0	7.6	−0.6	13.5
1,000 to 2,499	6.3	7.3	−1.0	12.3
2,500 to 4,999	6.1	7.5	−1.3	7.0
5,000 or More	5.4	5.6	−0.2	8.4

alternative measure. Evidently the regression fallacy illustrated in box 4.3 operates with powerful effect in the LRD data for the U.S. manufacturing sector.[17]

There is good reason to suspect that the regression fallacy operates with even greater effect in the longitudinal data sets used in the widely cited studies by Birch (1979, 1987) and the annual SBA reports. In particular, measurement error is almost certainly more serious in their data sets than in the LRD, a point we develop in the next section. Given their procedures for measuring firm size, the more serious measurement problems in their data suggest greater susceptibility to the regression fallacy.

In summary, the standard practice of measuring firm or plant size according to base-year employment leads to a regression fallacy, which in turn paints an overly favorable picture of the relative job growth performance of small employers. Our replication analysis with LRD data finds a substantial bias in favor of small businesses under the standard practice for measuring business size using base-year employment.

4.7 An Unsuitable Database

Still another weakness of many leading studies of the job creation process is their reliance on an unsuitable database: the Dun and Bradstreet Market Identifier (DMI) files. David Birch and associates use these data for their studies and, until recently, so did the SBA.[18] Although the Dun and Bradstreet database has many impressive attributes and represents an unparalleled source of information for many commercial purposes, it is not designed or maintained to maximize its usefulness as a tool for statistical analysis. Numerous studies

17. Brown et al. (1990) stress a different potential problem with the standard size measure. They argue that classifying new firms according to size in the entry year creates a bias because new firms often start small even when their intended scale of operations is large. This point clearly applies to new plants as well. However, a symmetrical point is that dying plants often contract and become small on their way toward exit. A careful reading of table 4.2 suggests that this latter effect dominates for manufacturing plants. Observe that among the smallest plants, the difference between the gross destruction rate based on current size and the gross destruction rate based on average size exceeds the corresponding difference for the gross creation rate. Observe, also, that the creation and destruction rates align more closely when comparing the current and average size measures than when comparing either of these measures with the Birch/SBA measure. This last observation indicates that the regression fallacy—not the birth problem stressed by Brown et al.—accounts for the striking contrast among panels in the table.

18. The SBA has contracted with the Bureau of the Census to longitudinally link the federal government's Standard Statistical Establishment List for the purpose of studying job creation and destruction behavior. See Census contract 61-93-41, "The Longitudinal Data Study."

have highlighted severe problems with the DMI files as a tool for measuring job creation and destruction or business births and deaths.[19]

For the purpose of investigating the job creation process, the DMI files suffer from two key problems. First, there is an enormous discrepancy between U.S. total employment as tabulated from the DMI files and the corresponding employment figures produced by the Bureau of Labor Statistics (BLS) or the Bureau of the Census. In 1986, for example, total employment tabulated from the DMI files exceeded the corresponding BLS and Census figures by 9 million persons.[20] In an economy with roughly 110 million employees, a discrepancy of this magnitude raises serious doubts about the accuracy of any statistical portrait generated from the DMI files. Furthermore, earlier research found that the most serious data problems in the DMI files involve younger and smaller businesses. This finding suggests that DMI-based claims about small business job creation should be interpreted with special caution.

Second, the DMI files do not accurately track business births and deaths or other important employment events. The U.S. General Accounting Office (GAO) has analyzed the accuracy of the DMI files in accounting for mass layoffs, with particular emphasis on layoffs due to plant closures. SBA provided GAO with a sample of mass layoffs and plant closures from the DMI files for the 1982–84 period.[21] The GAO study found that 81 percent of the mass layoff events in the DMI files were mistakenly identified. In reality, they represented some other event, such as a change in ownership structure, rather than a mass layoff or plant closure.

The DMI files also inaccurately identify plant births. A study by Birley (1984) compares three alternative sources of data for identifying new firms: the DMI file, the ES-202 data generated from administrative records maintained by state unemployment insurance agencies, and the telephone directory. She finds that the DMI files failed to identify 96 percent of the new firms found in the ES-202 data. Using a similar methodology, Aldrich et al. (1988) find that the DMI files missed 95 percent of apparently new businesses in the ES-202 data and 97 percent of those in the telephone directory.

In short, earlier research indicates that the DMI files are unsuitable for generating job creation and destruction figures. Identifying plant births and

19. See Armington and Odle (1982b), Birch and MacCracken (1983), Birley (1984), Howland (1988, chap. 2), Evans (1987b), Aldrich et al. (1988), and Small Business Administration (1983, 1987).

20. See Bureau of the Census (1986, p. 514).

21. The GAO defined a mass layoff as the dismissal of at least 20 percent of a plant's permanent workforce.