Preliminary and Incomplete
Do Not Cite Without Permission

Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes

By

Steven J. Davis
(*University of Chicago and NBER*)
John Haltiwanger
(*University of Maryland and NBER*)
Ron Jarmin, C.J. Krizan, Javier Miranda and Alfred Nucci
(*CES, U.S. Census Bureau*)[1]
Kristin Sandusky
(*LEHD, U.S. Census Bureau*)

March 2005

[1] This work is supported by the Kauffman Foundation and the U.S. Census Bureau. Any findings, opinions or conclusions contained is this paper are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

I. Introduction

The measurement of economic activity by federal statistical agencies focuses greater attention on larger, more mature business units. This data gathering strategy has two clear advantages. First, it yields greater accuracy in estimating the level of economic activity, whether "greater attention" takes the form of higher sampling probabilities or more careful auditing and editing. Second, it is easier to identify and promptly capture the activity of large, long-established business units. On both counts, the desire for a cost-effective approach to measuring the level of economic activity leads naturally to a focus on larger, more mature units.

There are, however, drawbacks to this data gathering strategy. When responses to shocks and new developments vary systematically with business size or age, a focus on larger and more mature units can yield less accurate, potentially misleading, measures of changes in economic activity. A simple example occurs when young and small business units are relatively sensitive to aggregate shocks. In this case, a cost-effective approach to estimating short-term growth rates can require over sampling younger and smaller business units, and there is tension between a sample design optimized for the level of activity and one optimized for the growth rate. Perhaps more important, the traditional focus on larger and more mature units limits our ability to measure and analyze the early life cycle dynamics of businesses and to evaluate theories of business formation, selection and growth.

In this paper, we report our initial efforts to remedy these drawbacks. We develop a preliminary version of an Integrated Longitudinal Business Database (ILBD) that combines administrative records and survey-based data for virtually all employer and nonemployer business units in the United States. In the process, we confront conceptual and practical issues that arise in measuring the importance and dynamic behavior of younger and smaller businesses. We also document some basic facts about younger and smaller businesses. In doing so, we exploit the ability of the ILBD to follow business transitions between employer and nonemployer status, and vice-versa. This feature of the ILBD opens a new arena for the study of business formation, early life-cycle dynamics, and market selection over time.

After characterizing the dimensions of the employer and nonemployer universes in the ILBD, we focus on roughly three dozen industries for which young and small businesses are relatively important. These industries contain 48 percent of all nonemployer businesses, 36 percent of nonemployer revenue, and 9 percent of employer revenue. Within this group of industries, nonemployer businesses account for 14 percent of total revenue, and young employers less than four years old account for another 17 percent.

The U.S. economy has more than 15 million nonemployer businesses as of 2000. Most are small and never transition to employer status or link to the employer universe in any way. Over a three-year horizon, five percent of the roughly 7 million nonemployer businesses in our selected industries migrate to the employer universe. The migrants stand out in some interesting respects. First, they are relatively large, accounting for 11 percent of nonemployer revenue three years prior to transition. Second, they grow considerably faster: Mean revenue growth for migrants in the year before transition is about 15 percentage points higher than for other nonemployers. Third, migrants account

for a sizable portion of revenue among young employers. Among employers in our selected industries that are less than four year old, we provisionally estimate that 17 percent of firms and 11 percent of revenues are generated by businesses that transitioned from nonemployer status in the previous three years.[2]

Nearly 12 percent of all employers in our selected industries link to nonemployer businesses in the previous seven years. In most instances, nonemployer status precedes employer status. In many other instances, the employer owns at least one establishment that predates the linked nonemployer business. These instances reflect an employer that absorbs a nonemployer business as well as separate employer and nonemployer businesses under common ownership.

As another step toward an integrated perspective on the dynamics of young and small businesses, we compare the growth and volatility patterns of employer and nonemployer businesses. We focus on revenue as an activity measure because it is available for all business units, employers and nonemployers alike. As in previous research, we find that net growth rates and their dispersions are highest for the youngest employers. We show that these two patterns also hold for nonemployers, and that net growth and volatility decline more rapidly at early ages among nonemployers. Holding age constant, the dispersion in growth rates is greater among nonemployers. Mean revenue growth rates exhibit no systematic relationship to size among employers, and they are $\cup$-shaped in size among nonemployers. The dispersion in revenue growth rates declines with size for both types of businesses.

The paper proceeds as follows. Section II discusses the construction of the ILBD, and Section III presents basic facts about employer and nonemployer businesses. Section IV offers a first look at linkages between the two business universes by considering the 2000 cross section of employers, and looking backwards for links to the nonemployer universe. Section V presents characterizes transition dynamics for particular cross sections of nonemployer and employer businesses. Section VI documents revenue growth and volatility patterns by business size and age. Section VII discusses next steps in our research program, and Section VIII offers concluding remarks.

II. Constructing an Integrated Longitudinal Business Database

A. Overview of Main Tasks and Previous Work

In terms of data development, our objective is a fully Integrated Longitudinal Business-level Database (ILBD) that includes all employers and nonemployers in the private non-farm economy. We construct a preliminary version of the ILBD for the years 1992 and 1994-2000, and we plan updates for more recent years. Key data for nonemployers are unavailable for 1993.

[2] As we explain below, these estimates are provisional in two respects and likely understate the true contribution of nonemployer-to-employer businesses to young employer firms. First, they rely on a conservative matching algorithm that understates nonemployer-to-employer transitions, perhaps by a substantial amount. Second, they reflect a crude adjustment for businesses that transition from nonemployer status outside our selected industries to employer status within our selected industries. Preliminary analysis suggests that this adjustment also understates the contribution of nonemployer-to-employer transitions. We will incorporate these transition cases into our analysis in the next draft.

From an analytical perspective, the presence or absence of employees is simply another business characteristic to be measured. From a database development perspective, however, integrating the Census Bureau's employer and nonemployer business is a major undertaking. One major task is the integration of the employer and nonemployer universes on a year-by-year basis. A second major task is the construction of longitudinal links for individual business. A third major task is the construction of contemporaneous and dynamic links between employers and nonemployers. Fortunately, we can build on previous work by Jarmin and Miranda (2003) to create the Longitudinal Business Database (LBD). The LBD covers all nonfarm private employers in the period from 1975 to 2001. We also build on previous efforts to construct longitudinal links within the nonemployer universe by Nucci and Boden (2003) and Boden and Nucci (2004).

B. Source Data for the ILBD

Census Bureau business registers draw on payroll tax records, corporate and individual income tax returns, applications for an Employer Identification Number (EIN), and various Census Bureau business surveys. The data available to the Census Bureau depend on the legal and tax status of a business and, in certain respects, on the size of the business and the number of its locations. For large corporations, routine data inputs include payroll records and particular items from corporate income tax returns, augmented by direct Census Bureau collections for multi-location companies. For sole proprietors, partnerships and single-location corporations with employees, routine data inputs include payroll records, certain items from income tax returns and periodic Census Bureau surveys such as the quinquennial Economic Census. For nonemployer businesses, routine data inputs derive mainly from income tax returns. Table 2.1 lists the most important administrative and survey sources for key variables in the employer and nonemployer universes.

Linking records from different sources requires common business identifiers. Employer businesses have unique Employer Identification Numbers (EINs) and other Census Bureau identifiers. Some nonemployer businesses also have an EIN, but most do not and instead are tracked by the person ID of the business owner, i.e., his or her Social Security Number.

To construct the ILBD, we must first ensure that administrative data from each universe are cleaned and ready for integration. On the employer side, this task has been largely accomplished in the work to create the LBD. The LBD contains high-quality longitudinal links for establishments, and for the most part it is easy to accurately determine establishment age.[3] Longitudinal establishment links are relatively straightforward to construct, because they are one to one, and because establishments typically have a clear physical location. Firms are more difficult to track over time, partly because firm-level links can be many to many. Work is underway at the Census Bureau to develop a rich set of longitudinal firm linkages. In the mean time, we define the age of a firm in the LBD as the age of its oldest establishment.

Longitudinal links are difficult to construct in the nonemployer universe for some of the

[3]See Jarmin and Miranda (2003). The main outstanding issue with respect to the LBD concerns the delayed identification of new establishments owned by certain multi-unit companies. We are developing algorithms to retime these births. The retiming issue pertains only to the recognition date of establishment birth, not the company-wide level of revenues or other measures of economic activity.

same reasons that they are difficult to construct in the employer universe. For example, when there is a change in the legal or tax status of a nonemployer business, its EIN or person ID can also change. (Person IDs do not change for individuals, but ownership changes can yield a change in the person ID associated with a business.) In these instances, it is not straightforward to maintain longitudinal links for nonemployer businesses using data items that are routinely included in the Census Bureau's administrative records systems. On the employer side, direct Census Bureau collections provide this additional information, but there is no ready equivalent on the nonemployer side. Hence, in the face of changes in legal and tax status, longitudinal links for nonemployers require matching algorithms based on data items such as business name, location and industry.

Integrating records from the employer and nonemployer business universes is further complicated by differences in the underlying administrative data and differences in the unit of analysis. The Census Bureau's Employer Business Register, which underlies the LBD, is a list of establishments (physical locations) maintained to serve as a mailing list for the Economic Census and as a frame for a variety of surveys. It relies heavily on administrative data and is augmented by direct Census Bureau collections.[4] Longitudinal linking is facilitated by establishment IDs (LBDNUMs and PPNs), EINs, enterprise IDs (Alphas), and business name and address information.

The Census Bureau's Nonemployer Business Register consists entirely of administrative data. The unit of analysis is a business entity, as recorded on a tax return. Our longitudinal links currently exploit EINs, person IDs, business name information, and geographic information. One complication arises when an individual taxpayer reports income for multiple nonemployer businesses. For example, multiple Schedule C forms can be attached to a single 1040 tax form. We deal with these cases by aggregating to the level of a single tax filing, which is associated with a unique SSN or EIN. We then create longitudinal links across years using SSNs and EINs in various combinations along with business name.[5]

In addition to the longitudinal links within each universe, we have created a preliminary set of links between the employer and nonemployer universes for our selected industries. These links make use of EINs, person IDs and business names found on records in both universes. For this purpose, we take advantage of several administrative data items. For example, when an individual applies for an EIN, he or she must fill out an SS-4 form for the IRS. This form includes the business name, the SSN of the business owner or chief officer and the EIN, all of which are included in the administrative records system in the Census Business Registers.

For this paper, we have taken a conservative approach in linking the nonemployer and employer universes. We have used all possible numerical ID matches and a conservative business name and geocode matching algorithm. Past efforts in developing longitudinal identifiers for the LBD has shown that linkages can be significantly improved by the use

[4] In order to track the establishment structure of multi-unit enterprises, the Census Bureau annually conducts the Company Organization Survey. This survey covers all large multi-unit companies and a sample of smaller ones. During an economic census, all establishments of multi-unit companies receive survey forms.

[5] Our current business name matching algorithm is conservative in the sense that it is unlikely to generate a spurious longitudinal link. We are currently developing more sophisticated matching algorithms that exploit name and address information.

of more sophisticated probabilistic matching algorithms that exploit all of the relevant available information and takes into account the reliability of the information. Given the conservative nature of our algorithm we are understating (perhaps significantly) the actual linkages between the two universes and thus, for example, understating the contribution of nonemployer businesses to formation of employer businesses.

III. The Employer and Nonemployer Business Universes

Table 2.2 provides summary statistics for the employer and nonemployer business universes in 2000. There are 15.5 million nonemployer businesses. Of these, roughly 13.4 million are person ID units (sole proprietorships without employees) and another 2 million are EIN units (corporations, partnerships and other nonemployer business entities with EINs).[6] There are another 5.4 million employer businesses. Of these, 182 thousand are multi-unit (MU) enterprises with multiple establishments, and the rest are single-unit (SU) businesses. While comparatively small in number, multi-unit enterprises account for 61% of aggregate U.S. business revenue. Nonemployer business units account for 4% of aggregate revenue, and single-unit employers account for 35%.

Given the sheer size of the Census Bureau business registers and some complex issues of measurement, we focus on a selected set of industries for this paper. We chose industries with large numbers and relatively high revenue shares for young and small businesses. Links between employers and nonemployers are likely to be relatively important for these industries. We avoided industries with complex measurement issues related to financial holding companies, tax shelters, and special purpose financial entities. These aspects of corporate organizational structure are interesting, but they are not the focus of our data integration and analysis efforts.

Our analysis period overlaps with the transition from SIC to NAICS industry classifications, and the SIC-NAICS crosswalk is a many-to-many mapping. Our nonemployer data files contain 3-digit SIC codes prior to 1997 and 4-digit NAICS codes thereafter. The employer data files contain codes for both classifications from 1997 to 2000. Accordingly, we proceed as follows. For many exercises, we look backwards for businesses in selected 4-digit NAICS industries. For other exercises, we look forward from a year prior to the NAICS changeover at businesses in selected 3-digit SIC codes.

Table 2.3 provides summary information for our selected industries. Legal Services has the largest number of employer businesses, almost 150 thousand. Other Personal Services has the largest number of nonemployer businesses, more than 800 thousand. The highest-revenue industry for employers is Gasoline Stations at 187 billion dollars.[7] The highest-revenue industry for nonemployers is Real Estate Agents and Brokers with almost 23 billion dollars. In terms of employment and payroll, Legal Services is the

[6] The distinction between person ID and EIN units involves some complexities. For example, a sole proprietor with no payroll but positive receipts, and who has applied for an EIN, may not appear only in the person ID portion of the Nonemployer Business Register. In addition, these businesses may also appear in the Employer Business Register, as it is maintained by the Census Bureau. We assign all zero-payroll units to the nonemployer universe, even if they reside in the Employer Business Register.

[7] High revenue in Gasoline Stations mainly reflects the cost of gasoline. The administrative data in the Census business registers typically does not include information on gross margins or material costs. The data are included in the Economic Censuses and various annual surveys.

largest industry with more than one million workers and more than 58 billion dollars in payroll.

Table 2.4 provides information about industry shares of total revenues and the relative size of employer and nonemployer segments within industries. Nonemployer revenue shares range widely, and they exceed 30% for a dozen of our selected industries. At the upper end, nonemployers account for more than two-thirds of revenue in Independent Artists, Writers and Performers. At the lower end, nonemployers account for a mere 0.24% of revenue in Software Publishers, even though 25% of all businesses in this industry are nonemployers. Although extreme, the basic pattern in this industry is not uncommon; it reflects the enormously skewed size distribution of activity in many industries.

Figures 2.1 through 2.4 provide information about the age and size distribution of businesses in these selected industries relative to the entire economy. Business age here is defined as the first time the entity is seen in its respective universe (for employer firms this is the age of the firm's oldest establishment). We construct age categories in a consistent manner in both universes: 1, 2, 3, 4, 5, 6-7, 8+. We combine age classes 6 and 7 since the nonemployer data are missing for 1993 so that looking backwards from 2000 we must combine the 1993 and 1994 entering cohort.

When comparing the size distributions across the two universes we use revenue size classes as this is the only comparable activity measure across the two universes. The revenue cutoffs we use (in thousands of year 2000 dollars) for revenue size classes are: less than 3, 3-6, 6-9,9-12,12-30, 30-90,90-180,180-360, 360-600, 600-1200, 1200-3000, 3000+. The lowest size classes are, as will soon become clear, somewhat less relevant for the employer universe but we include them here since they are quite relevant for the nonemployer universe. Likewise, the largest size class is somewhat less relevant for the nonemployer universe but we include since it is so important for the employer universe.

In terms of age, even for our selected industries, almost 70 percent of revenue and more than 40 percent of employer firms is accounted for by firms 8 or more years old. For nonemployer firms, the share accounted for by firms 8 or more years older is much smaller. For nonemployers in particular, the firm share and revenue share distributions are u-shaped with very young firms accounting for a larger share than say 5 year old businesses. In terms of the size distribution, most nonemployer businesses are very small earning less than $3K in annual revenue but not surprisingly such firms account for a very small fraction of revenue for nonemployer firms. For employer firms, almost seventy percent of revenue is account for by firms with more than 3 million dollars in annual revenue but such firms account for less than five percent of the share of employer firms. Most employer firms are in the middle size classes depicted in Figure 2.2. Indeed it is striking the number of employer firms that have between 30K and 360K of annual revenue.

The tremendous variation in size across nonemployer and employer businesses exhibited in Figure 2.2 suggests appropriate caution in drawing inferences about the behavior of "small and young" businesses per se. The size distribution variation alone reminds us that many nonemployer businesses are extremely small reflecting some secondary or supplemental means of generating income. Analyzing the dynamics of such businesses alongside much large employer and nonemployer businesses is a challenge for this effort to construct and analyze the ILBD. In what follows, we often report results on both a

firm and revenue weighted basis. The former by construction provides more insights about the behavior of the very small (and more prevalent) businesses while the latter provides more insights into the contribution of the larger businesses.

In terms of specific industries, Figures 2.3 and 2.4 depict the revenue and firm shares of young and small businesses. The upper panel of Figure 2.3 shows that industries with a high share of young firms are also industries where the young firms account for a large share of revenue. The same pattern holds for the employer businesses in the lower panel of Figure 2.3 but most industries are far to the right of the 45 degree line. In terms of the size distribution, the upper panel of Figure 2.4 shows the positive relationship between firm and revenue share is evident for the nonemployer businesses in Figure 2.4 but with all industries far to the right of the 45 degree line. For employer businesses, the revenue share by small businesses is so small that the relationship is essentially on the horizontal axis.

For specific industries, the patterns vary substantially across the nonemployer and employer universes. For example, for Software publishers (5112), young businesses account for a very large share of firms and revenue for nonemployer businesses but relatively a much smaller share in the employer universe. Moreover, in terms of the size distribution, while about 25 percent of software firms are small in the employer universe, such firms account for less than 1 percent of revenue in this industry.

III. Backward-Looking Links from Employers to Nonemployers

A key objective is to understand the linkages between records in the employer and nonemployer business registers and the businesses they represent. As a first step, we start with the 2000 LBD and limit the analysis to our selected group of industries. From both an economic measurement and research perspective, we want to understand the longitudinal characteristics of businesses in the LBD in 2000. For instance we want know things like how old the business is, or whether it originated as a nonemployer business. Thus, we match the 2000 LBD for our selected industries to the Nonemployer Business Register for all available years between 1992 and 2000. Since the LBD is an establishment file and the Nonemployer Business Register is a tax return based file, we need to role the observations in each dataset up to the level of the matching fields in common across them. These are EIN, person ID, Business Name and geocodes. The LBD records, aggregated to the appropriate level, are linked to each year of available nonemployer data. Once links are obtained, LBD records are then aggregated to the enterprise level since it is not possible, at this point, to tell which specific establishments, if any, are associated with a linkage to the nonemployer data.

The result is an LBD based enterprise level dataset with linkages to the nonemployer data and additional variables that describe the nature of the nonemployer records to which the enterprise is linked. Note that the unit of observation here is an enterprise operating within one of our selected 3-digit SIC industries. That is, firms operating in more than one of these industries will have multiple observations in these data. As noted in section 2, the matching algorithm we use for this paper is conservative so we are understating the likely linkages between the two universes.

Table 3.1 summarizes the age and size distributions of employer businesses from the LBD in 2000 and, in turn, the nature of the backward linkages to nonemployer businesses. Familiar patterns of business age and size distributions are present. That is,

for any cross section, the number of active businesses declines with both age and size, but the bulk of activity – measured here by payroll – is concentrated in old and large businesses.

Our main focus in this table, however, is to see how a cross section of employer businesses is linked to nonemployer businesses. When we think of business formation and growth, the typical pattern of nonemployer to employer linkages we would expect to see is where a business starts as a nonemployer, grows and eventually transitions to the employer universe where it grows further still. In our work to date, however, we take a more agnostic view and simply search for any and all linkages between our cross section of employer businesses and the nonemployer universe.[8]

The fifth column in Table 3.1 gives the number of employer businesses that link to one or more nonemployer businesses between the years 1992 to 2000 (excluding 1993). We break these preliminary linkages out by the size and age of the employer business in the 2000 LBD. Note that for this purpose, our size definitions are based upon the number of employees. In turn, we show the share of firms and share of linked nonemployer receipts that are in the various size and age classes. To measure the share of nonemployer receipts we use the deflated average annual receipts of linked nonemployer records for the two years prior to the link.

Of the 2.3 million employer businesses in our industries in 2000, about 266 thousand have links to the nonemployer universe in the prior 8 years. A large share, (30.4%), of all employer firms with links to the nonemployer data are aged eight years old or older. Second, a very large share (75.4%) of firms with links are small (fewer than 5 employees). In terms of the share of the nonemployer receipts, the very small employer businesses account for 46.73 percent of the nonemployer receipts with links and the more mature businesses account for a large fraction of nonemployer receipts as well. Thus, it is interesting that the links between the employer and nonemployer businesses are dominated in the employer universe by very small but also more mature businesses.

The high portion of linkages to the nonemployer universe accounted for by older employer businesses suggests that transitions of young growing businesses is not the only way in which nonemployers enter the employer universe. We examine the timing of these linkages in more detail in Figure 3.1. We know the age of employer businesses from the LBD. We construct an age measure for a nonemployer business based on the first year we observe it in the Nonemployer Business Register. We can observe employer businesses in the LBD all the way back to 1975, but our nonemployer data go back only to 1992. Thus, many nonemployer businesses observed in 1992 will actually by much older (just as many employer businesses observed in 1975 will be much older). In results not reported here, we dropped all cases where the nonemployer business was observed in 1992 and found no qualitative differences in the results.

Figure 3.1 provides a comparison of the age distributions of linked employer and nonemployer businesses. Only linked cases are used in the figure which shows the frequency of the computed difference in the age of employer and nonemployer records that link. The modal outcome is positive one indicating that most links occur where the nonemployer business is one year older than the employer business. Moreover, the

[8] The at risk group for this analysis are all employer businesses in 2000 with positive payroll in our selected industries. In looking backwards for links we do not require that the link to the nonemployer business be in the same industry. The link is via ID and/or business name.

distribution over the range for which we don't have left censoring problems is skewed to the right. The fact that the mode and most observations have a positive age difference suggests that most of the transitions are "standard" in the sense that the nonemployer business exists prior to the employer business. However, there are a substantial number of linkages where the employer business is older than the nonemployer business. For these cases in particular, it appears that the linkage describes an employer business absorbing a nonemployer business.

IV. Transitions between Nonemployer and Employer Status

In the prior section, we undertook a preliminary examination of the linkages between the employer and nonemployer business universe. The primary results are that there are extensive and complex linkages between the two business universes. The extent of the linkages suggests that integrating the employer and nonemployer business universes is critical for understanding business dynamics and especially the dynamics of young and small businesses. However, the nature of the linkages is complex with nonemployer businesses transiting to employer status and nonemployer businesses being absorbed by older employer businesses. In this section, we continue this exploration into the linkages and transitions between the two business universes by conducting a forward-looking analysis of transitions for both business universes. In particular, we examine the 1994 nonemployer business universe and explore the status of all such 1994 businesses three years later.[9] In turn, we conduct a parallel analysis for the 1994 employer business population.

Figure 4.1 depicts the three year transition dynamics for the 1994 population of nonemployer businesses.[10] The specific population we use are the nonemployer businesses with positive receipts in 1994 in our selected industries. Linkages over time are based upon person ID and EIN matches over time as well as business name and geocode matches.[11]

We classify the possible outcomes into four categories. One category is business exits which in this setting reflects businesses with positive receipts in the nonemployer universe in 1994 and zero receipts in the nonemployer universe and zero payroll in the employer universe three years later (in 1997). A second category is nonemployer business who transit completely to the employer universe in that they had positive receipts in the nonemployer universe in 1994, zero receipts in the nonemployer universe in 1997 but positive payroll in the employer universe in 1997. A third category is continuing nonemployer businesses with no links to the nonemployer universe. A fourth category is another form of linkages or transitions which are nonemployer businesses

[9] We have also computed these transition shares over a six year horizon and the basic patterns are very similar, although the magnitudes change in expected directions (e.g., the share of activity accounted for by exits on both a firm and revenue basis increases substantially).

[10] We have also computed these transition shares over a six year horizon and the basic patterns are very similar, although the magnitudes change in expected directions (e.g., the share of activity accounted for by exits on both a firm and revenue basis increases substantially).

[11] As in the prior section, we don't require the forward link to be in the same industry. The categories we examine in terms of continuers and exits as well as links are not dependent on being in the same industry. In the subsequent analysis of employer to nonemployer transitions, we follow a similar protocol. That is, the at risk group are employer businesses in 1994 with positive payroll in our selected industries. Forward links are not dependent upon being in the same industry.

who have positive receipts in the nonemployer universe in 1997 but also have positive payroll in 1997 in the employer universe. These latter businesses might result from a number of different factors. For one, we may be capturing the transition year where the business operates in both universes. For another, this may reflect a business owner who owns multiple businesses some of which have zero employees and others have positive employees.

About half of the nonemployer businesses and about 60 percent of the revenue in 1994 are continuing businesses without any links to employer businesses in 1997. Almost forty percent of the businesses and about twenty-five percent of the revenue are associated with exits. About three percent of businesses but about seven percent of revenue in 1994 are complete transitions to the employer universe. While three percent is not a large share per se, since there are over five million nonemployer businesses in our selected industries, this reflects more than 220,000 businesses in this "complete transition" category. The continuing nonemployer businesses that have some link to the employer universe in 1997 constitute about 2 percent of firms and 4 percent of revenue in 1994.

Figure 4.2 reports the analogous transitions with person ID firms and EIN firms broken out separately. For both exits and continuing nonemployer businesses, the patterns are similar although exits are even more prevalent for the EIN firms. We find that complete transitions are more prevalent for the EIN firms and links where the business exists in both universes simultaneously are more prevalent in the person ID universe.

These basic transition patterns raise more questions than they answer. Basic questions include: what are the factors that determine the patterns that we observe? and/or even more generally what are the characteristics of the businesses in each of the transition categories? While it is beyond the current paper to answer these questions fully, Figure 4.3 takes a basic step in answering these questions by providing insight into the growth dynamics of the businesses in each of the categories prior to any transition that may have occurred. Specifically, we depict the growth rate distribution in 1995-96 for different types of businesses. The upper left panel depicts the growth rate distribution for all 1994 nonemployer businesses with links to the employer universe in 1997 and who had positive activity in the nonemployer universe in 1996. This panel combines the second and fourth categories in Figures 4.1 and 4.2 but focuses on the year 1995-96. The top right panel depicts the growth rate distribution for the complete transition cases (this is the second category in Figures 4.1 and 4.2). For these cases, the growth rate distribution reflects the year just prior to making the complete transition. The lower left panel depicts the growth rate distribution for the businesses with links but who also remain in the nonemployer universe (this is the fourth category in Figures 4.1 and 4.2) . Finally, the lower right panel shows businesses that did not transit, including the continuing businesses without links and the exiting businesses (this panel combines the first and third categories from Figures 4.1 and 4.2).

The transiting businesses have a substantially higher growth rate than those without any links. The mean growth rate for all transiting businesses is 0.07, the mean growth rate for complete transitions is 0.02, the mean growth for other transitions is 0.12 and the mean growth for the businesses with no links is -0.10.[12] The transiting businesses also have

[12] Even if we only consider continuing nonemployer businesses, non-transiting continuing nonemployers have a lower growth rate (-0.02).

less dispersion in growth rates. The interquartile range for all transiting businesses is 0.33, for complete transitions it is 0.29, for other transitions it is 0.36 and for businesses without links it is 0.42.

Another summary measure of volatility of growth rates that is useful to consider in this context is the excess reallocation rate which follows the concepts developed by Davis, Haltiwanger and Schuh (1996) (hereafter DHS). In this context, the measure of activity is not employment but rather revenue but the basic idea is the same.[13] The excess reallocation rate for all transitions is 25.6, for complete transitions it is 30.4, for other transitions it is 21.2 and for businesses without links it is 32.8.

This analysis shows that transitions have both a higher growth rate and less volatility of growth relative to their counterparts without links. These patterns make intuitive sense and, in turn, help us better understand the nature of these transition dynamics. The high net growth rate for the transitions suggests that businesses prior to making the transition tend to have disproportionately high growth, and, interestingly relatively low volatility.

We now briefly examine the contribution of these transitions to the employer universe entrants. We know from the analysis in the prior section that transitions to the employer universe involve both nonemployer businesses becoming new employer businesses but also nonemployer businesses being absorbed by older employer businesses. To gauge the contribution from our 1994-97 transitions, we conduct two related exercises. First, we simply consider the share of firms and revenue in the 1997 employer universe accounted for by the transition cases. These shares are computed for each 3-digit SIC industry and reported in the top panel of Figure 4.4. For most industries, the share of firms is less than 15 percent but in some industries the share of revenue accounted for by employer businesses that have a link to the 1994-97 transitions is around 50 percent. In particular, in both the Trucking and Courier and Business Service industries, the share of revenue is around 50 percent. In such industries, it would appear that the transiting nonemployer businesses play a very large role. However, in these and other industries it may be that it is the employer business absorbing the nonemployer business.

To gauge the contribution of the transitions to employer exits, we change our focus to employer businesses less than three years old in the 1997 employer universe. For all of those young/entering businesses, we compute the share of firms and the share of revenue associated with the subset of those businesses that represent transitions from the nonemployer universe. Figure 4.4 shows the shares by industry. The contribution to

[13] For each business, the growth rate is measured as the change in revenue from time t-1 to t divided by the average revenue for the business in t-1 and t. This growth rate concept is discussed in detail in DHS but has desirable properties like the log first difference growth rate measure in that it is symmetric for positive and negative changes but has the advantage relative to the log first difference in that it can accommodate startups and shutdowns. Moreover, this measure is a second order approximation of the log first difference growth measure. A gross revenue expansion rate is calculated as the weighted average of the growth rates for all expanding businesses (including startups) where the weights are the average revenue of the business in periods t-1 and t. A gross revenue contraction rate is calculated as the weighted average of the growth rate for all contracting businesses (including those shutting down). The sum of the expansion rate and the contraction rate is the total reallocation rate. The latter is a useful summary measure of volatility (and indeed is an absolute deviation measure of dispersion centered around zero) but this measure of volatility will by construction be positively related with the magnitude of the absolute value of the net growth rate. To overcome this limitation, we use an "excess" reallocation rate which is the total reallocation rate less the absolute value of the net growth rate (see DHS for more discussion of these concepts).

entry in terms of firms and/or revenue is small but substantial especially in selected industries. For eating and drinking establishments (580), insurance brokers (641) and personal service businesses (729) the share of revenue is almost 20 percent.

In both of these exercises, we are understating the contribution of the transitions to activity in the employer universe because of an interesting and challenging data issue. It turns out that about 1/3 of our 1994-97 nonemployer to employer transitions when entering or linking to the employer universe switch to an industry outside of our 3-digit list of industries. Much of this switching is from some general miscellaneous industry to a better defined industry so in many cases this switching may be in fact an improved and more precise industry code. While we can eventually track these industry switchers since we have access to the full universe files, for this paper where we focus on selected industries we can take this point no further. For now, we note that the shares in Figures 4.4 are understated given that roughly 1/3 of the total nonemployer to employer transitions are not included in this analysis due to industry switching. Using this share of industry switchers as a crude adjustment factor suggests that the actual contribution of the nonemployer transitions to the employer universe in 1994 should be roughly 1.5 times the shares reported in Figure 4.4. Using this adjustment factor with the shares reported in Figure 4.4 yields provisional estimates that suggest that 17 percent of the number of young employers and 11 percent of the revenue of young employers are accounted for by transiting nonemployer businesses.

To conclude this section, we examine transitions going the other way – i.e., we begin with the universe of 1994 employer businesses and examine their outcomes at a three year horizon. Figure 4.5 reports the results for these transition dynamics that are defined in an analogous manner. The primary difference here is that for the employer universe, positive activity is defined in terms of positive payroll so the universe at risk is employer businesses with positive payroll in 1994. In defining categories, we use the information in 1997 about payroll in the employer universe and receipts in the nonemployer universe. Exits are businesses with zero (employer) payroll and zero (nonemployer) receipts in 1997 while continuers without links have positive payroll in 1997. Complete transitions in this case are businesses with positive receipts in 1997 in the nonemployer universe but zero payroll. Other transitions are employer businesses that are continuing businesses but with a link to a nonemployer business with positive receipts in 1997.

In examining these transitions, by far the most important category are continuing employers without links to the nonemployer universe. The second most important category are exits from the employer universe. While this qualitative pattern matches the analogous transitions for nonemployers reported in Figure 4.1, it is notable that exits for employer businesses are much less likely than exits for nonemployers.

In terms of employer businesses with links to the nonemployer universe, a small share are continuing employers with positive 1997 receipts in the nonemployer universe. The continuers with links with positive receipts in part reflect individuals who own both an employer and a nonemployer firm. The share of businesses (on a firm basis or revenue) basis that transit completely to the nonemployer universe is quite small. For example, the complete transitions to the nonemployer universe in 1997 account for about 3 percent of firms and 1.5 percent of revenue in the employer universe in 1994. Comparing the analogous category in Figure 4.1 this category of transitions from employer to nonemployer is much smaller especially in terms of number of businesses. In this regard,

it is important to emphasize that the number of nonemployer businesses is about 3 times the number of employer businesses so three percent of the number of employer businesses is about 1/3 of three percent of nonemployer businesses.

V. Business Revenue Growth and Volatility by Age and Size

The prior sections make clear there are a rich set of linkages between the employer and the nonemployer universes. The nature of the linkages is complex. Some reflect transitions such as a nonemployer business becoming sufficiently successful that it transits to employer status (i.e., the business hires workers). Others reflect existing employer businesses absorbing nonemployer businesses. Still others apparently reflect individuals or entities that operate simultaneously (or at least frequently) in both the employer and the nonemployer universe. Given our focus on the dynamics of young and small businesses, these rich connections imply that the study of only employer businesses and/or only nonemployer businesses separately may yield some misleading characterizations of dynamics – particularly dynamics with respect to entry and exit and more generally the role of business age. Our long run objective is to integrate the employer and nonemployer business universes to provide a comprehensive data infrastructure that will enable us to look at the core issues of young and small business growth taking the linkages between the employer and nonemployer universes into account. However, as should also now be clear from the analysis above, the linkages are sufficiently complex that constructing an integrated, longitudinal business database incorporating employer and nonemployer businesses is a major data infrastructure undertaking. Our goal in this paper is to explore the connections and relationships between the two universes to help us to this long run objective. With this in mind, in this section we undertake a comparison of the growth rate dynamics we observe in each of the universes independently. Our limited objective here is to compare and contrast the dynamics observed in each universe, knowing that a full understanding of these dynamics awaits the fully integrated data infrastructure.

Specifically, in this section, we examine the patterns of net growth and growth rate volatility (using the excess reallocation concept) for both the employer and nonemployer universes by measures of business age and business size. As before, when looking backwards and using age and size as the business characteristics, we focus on the reference year 2000 (really the growth between 1999-2000 as this is the most recent year we can measure growth and volatility in a consistent manner in both business universes.

Figure 5.1 depicts the pattern of net revenue growth rates for the 1999-2000 year by age class and Figure 5.2 depicts the pattern of volatility (excess revenue reallocation rates) by age class for both the employer and nonemployer universes. Since we have more years we can construct these statistics for the employer universe, we include the 2000 statistics but also the time series average of the same statistics for the years 1994-2000 for the employer universe. Note that all of the statistics reported in this section are based on activity-weighted growth rates as described in section 4. This implies that at least within a category (age or size class) businesses that are larger are given a larger weight.

For both employer and nonemployer businesses, net growth and volatility decline with business age. However, the pace of the decline of net growth with age is much sharper with nonemployer businesses suggesting that the first few years of an employer business are especially critical in terms of adjusting to the size it operates in while still young. In

contrast, volatility of nonemployer businesses declines more sharply with business age and volatility of nonemployer businesses is substantially higher for all age groups. In addition, volatility of revenue for employer businesses is mildly hump shaped.

The magnitudes of the excess reallocation rates are very large for all age groups but especially for young, nonemployer businesses. For example, for 1 year old nonemployer businesses, the excess revenue reallocation rate exceeds 80 percent. Since this measure abstracts from the net change for this category, this measure suggests that more than 80 percent of the revenue for very young nonemployers is reallocated within a year. Even for mature, employer businesses we observe relatively high excess revenue reallocation rates with rates in excess of 30 percent. The churning of business activity across firms is a ubiquitous feature of the U.S. economy and it is striking that it is especially high amongst young businesses in general and nonemployer young businesses in particular.

Figures 5.3 and 5.4 show the analogous patterns for revenue size classes. The net growth relationship with size is quite different across employers and nonemployers. For nonemployers, net growth is u-shaped with respect to size. In contrast, for employers, net growth increases with size in 2000 but is hump shaped for the average. For the very smallest size classes this apparent volatile pattern over time may not be especially interesting since the share of employer businesses in the smallest size classes is very small but it is also interesting that the pattern is not systematic for the larger size classes.

In terms of volatility, Figure 5.4 shows that the volatility declines with business size for both employers and nonemployers. For the very small businesses, volatility is actually greater for employers but once businesses have at least $180K in annual revenue, this relationship is reversed (i.e., nonemployers are more volatile). Interesting volatility for employer and nonemployer businesses is about the same for the largest businesses.

Finally, as a robustness check, Figure 5.5 reports analogous net and volatility measures for payroll for the employer universe only and by business age. We find that the patterns for payroll based measures are similar to those for revenue (in Figures 5.1 and 5.2) but less noisy which makes sense as the payroll numbers are more reliable in the employer universe. In addition, there is no evidence of the mild hump shape in the volatility measure for younger ages as there is for revenue. It may be of interest to explore why the payroll and revenue volatility patterns differ.

Our simple comparisons of growth dynamics between the employer and nonemployer universes are interesting, but they neglect business transitions between the two universes, as do almost all previous studies of business dynamics. The complex pattern of linkages between the two universes that this paper begins to explore raises a variety of issues regarding the measurement and analysis of business formation, growth and volatility, especially for younger and smaller businesses. Our initial foray into this territory suggests that the Integrated Longitudinal Business Database will prove extremely useful in developing a clearer, more comprehensive picture of business dynamics.

VI. Where Do We Go From Here?

In this paper, we have described the broad features of the employer and the nonemployer

firm universes and the relation between them over time. Many database issues await us, both for the employer and nonemployer files. In addition, the findings presented in this paper – and the many caveats surrounding them – while preliminary already allow us to see the many research questions that arise from access to these firm data. In this section we explore two related topics: (i) an overview of some of the core research questions ahead of us and (ii) the data infrastructure and measurement challenges we face in pursuing this research agenda.

Access to firm-level information -- over time – for both employers and nonemployers allows us to address the more general question (s) how the two universes – treated together and netting out transfer between them -- affects research findings on firm dynamics in one of these firm populations, conducted in isolation from the other (Blanchflower, 1998;Caves, 1998). Exploring firm dynamics with such integrated data provides a rich new perspective relative to the existing literature since most studies of firm population dynamics are studies of employers firms. The literature on nonemployer firm dynamics is sparser and is plagued by limited sample size, cross sectional data structure or some combination of both. The comprehensive nonemployer firm data that are part of this effort are of considerable interest in their own right. Using these data alone will provide an unprecedented the first to allow us to ask questions not only about the behavior of the nonemployer over time – with robust samples, even censuses – but the demographic and economic characteristics of the owners and the firms. These integrated employer and nonemployer firm data allow us to address a variety of research questions. For example, we can study the early "beginnings" of firms both for employer and nonemployer firms. In addition, we can study the dynamics of the self-employed in a unique fashion including self-employed who employ workers versus self-employed who are non-employers.

To be able to explore such issues, there remain many data base issues not addressed – or partially so – in our efforts thus far. We may view the issues as either intrinsic to one of the two firm populations – employer or nonemployer – or encountered when bringing the two together. Some issues are common to employer and nonemployer data files: the linkages of the businesses over time within and between the two universes; the conversion from SIC industry codes to NAICS ones following the 1997 economic censuses; reliability of geography and industry classification and data where source data are administrative. Others are specific to individual datasets.

Data issues unique to the employer data base are primarily related to longitudinal linkages and to the breakout of multiunit businesses as well as related spikes in changes in geographic and industry codes in economic censuses. In terms of longitudinal linkages, while considerable progress has been made in developing and maintaining longitudinal establishment identifiers, development of firm-level longitudinal identifiers is an open area for research and development. By construction, firm-level identifiers change when a business undergoes any number of types of reorganization including changing legal form of organization or merger and acquisition activity. In the analysis above we have confronted this issue in the development of firm level age measures which we define as the age of the oldest establishment but we currently define entry and exit of firms based upon changes in the firm level identifiers. Another issue surrounds the treatment of multiunit businesses. Multiunit businesses above a size threshold are

included in the Census Company Organization Survey (COS). However, the list of such businesses is drawn from the prior economic census. These procedures imply that single units that become multiunits are typically not detected until the next economic census. Further, new establishments from small multiunits not covered by the COS are detected only at the economic censuses. In both cases, measures of the economic activity of these new establishments is captured in the employer universe in non-Census years with the original business units but the count of establishments as well as potentially the industrial classification and geographic classification of activity are mismeasured. Moreover, at the time of the Economic Censuses, these new establishments are detected so that there is a spike in the number of new establishments with associated spikes in the amount of activity that changes industry and geographic location.

Data issues unique to the nonemployer firm universes are driven by the relatively recent availability and development of annual nonemployer firm data at Census. Further, the nonemployer firm data and industry/geography classification are mostly administrative. Hence we are less clear about our understanding of the data editing procedures over the years and how this affects classification and data items. The construction of a longitudinal file of nonemployer firms raises further challenges as we are at an early stage of development here and several components of this development remain exploratory. This includes improving the longitudinal linkages via name and address matching (both cross sectionally and longitudinally), the treatment of joint returns for sole proprietorships (where there are two firm id's, filer and spouse), and the reliability of employer identification information for nonemployer sole proprietors, and the reliability of filer id's on the employer database.

Data issues that are common to both employer and employer universes include (i) improving the quality of the linkages between the employer and nonemployer universes; (ii) developing comprehensive SIC-NAICS concordance (s); and (iii) underlying all of this are data reliability issues for measures of industry, geography, and economic activity. One challenge here in that latter regard is that there are large portions of the two universes that rely almost exclusively on administrative sources. It is also worth noting that the SIC-NAICS conversion is more of a challenge in the nonemployer universe since the nonemployer industry codes are inherently less detailed and precise. Also, the reliability of the revenue measures is further complicated in both files as they are subject to (changing) business income tax return conditions in addition to issues associated with the item itself (see Table 2.1).

Bringing together the two firm universes – employer and nonemployer – raises two additional issues: 1) netting out those units that are likely duplicates (e.g., nonemployer observations that are really employer units) for the same year vs. those units that legitimately operate in both universes simultaneously; 2) identifying those units in one universe that appear in the other universe at a subsequent/prior period. In the above analysis, we have used relatively conservative rules for linking the two universes. Refined matching techniques may yield a significant increase in the number and types of matches and links across the universes.

Studies of the universes combined – in the sense described above and where we are going – face further challenges. One, the standards for industry/geography classification may

differ across these two universes and thus, inhibit out ability to isolate industries/region of interest in our analyses. This measurement challenge is related to the fact that the employer universe includes information drawn from survey/economic census data collections on these fields while the nonemployer universe does not.. In addition, additional measurement challenges relates to the fact that the employer universe typically of interest is primarily one of establishments (remembering we can always summarize these establishment to their tax report or ultimate owner) while the nonemployer universe consists of tax returns units.

One additional important direction both conceptually and in terms of measurement will be the integration of *employee* data into these business universes. Using the longitudinal matched employer-employee data from the LEHD program at Census, the universe of employees can be integrated into these files along with related information about the demographics of these employees.[14] Integration of the employer, nonemployer, demographic and employee data provides an unprecedented opportunity to study firm, job and worker dynamics in an integrated and comprehensive fashion. The integration of the nonemployer data with the employee data open up the rich possibility of being able to understand the role of nonemployer businesses that are the primary vs. the secondary sources of income for the individual. Moreover, for the questions about where employer businesses come from, knowing about the demographics of the founders as well as knowing the history as both a nonemployer and as an employee of the founders of a employer business has rich possibilities.

VI. Concluding Remarks

To conclude this paper, we begin with a simple analogy that helps in thinking about the role of nonemployer and employer businesses and their respective connections. Consider the life cycle of dandelions. There are enormous numbers of dandelion spores, the vast majority of which never become dandelions. However, a small share become dandelions and it is basic fact that no dandelion becomes a dandelion without first being a dandelion spore. So while most dandelion spores will never be dandelions, all dandelions come from dandelion spores.

For our purposes, we think of the nonemployer businesses as analogous to the dandelion spores. Most are very young and small and will never become employer businesses. However, a small but important share become employer businesses. The open question here is the contribution of such transiting nonemployer businesses to the employer universe. We don't believe the integration of the employer and nonemployer universes is sufficiently developed to provide a comprehensive answer to this question. Our preliminary evidence suggests the contribution is nontrivial. While our analysis is too preliminary to be definitive, our evidence suggests this dandelion analogy breaks down on one key point – many employer businesses apparently have no history as nonemployer businesses. Thus, while all dandelions come from dandelion spores, not all employer businesses come from nonemployer businesses. Still, our preliminary findings suggest

[14] Another obvious direction for data integration is to integrate the Characteristics of Business Owners (CB0) and Survey of Business Owners (SBO) surveys into the ILBD. Holmes and Schmitz (1995), amongst others, have shown the rich analysis that can be conducted with the CBO. Integrating the CBO into the ILBD offers an additional rich set of possibilities.

that the contribution of the nonemployers themselves to economic activity is substantial in its own right, such nonemployer businesses exhibit rich dynamics and the contribution of transiting businesses is substantial enough that exploring these connections further is clearly a rich area for future research.

Our ultimate goal is to construct and analyze an Integrated Longitudinal Business Database based upon an integration both cross sectionally and longitudinally of the employer and nonemployer business universes. This paper represents small but important steps in this direction. The findings to date quantify the importance and the measurement difficulties that must be confronted in constructing the integrated business database. The complex nature of the linkages between the employer and the nonemployer universes are the biggest challenge we face in working towards this objective.

References

Blanchflower, David G. 1998. "Self-employment in OECD Countries." Paper presented at Canadian International Labor Network (CILN) Conference, Burlington Ontario, September 24th-26th 1998.

Boden, Richard and Alfred Nucci. 2004. "Business Dynamics among the Smallest and Newest of Businesses: Preliminary Analyses of Longitudinally Matched Nonemployer and Business Register Data." Annual Meetings, Eastern Economic Association, Washington, D.C. March , 2004.

Caves, Richard E. 1998. "Industrial Organization and New Findings on Turnover and Mobility." Journal of Economic Literature. Vol XXXVI (December, 1998), pp. 1947-1982.

Davis, Steven J. and John Haltiwanger, "Gross Job Flows" in *Handbook of Labor Economics* (Orley Ashenfelter and David Card, eds.), North Holland, 1999.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, MIT Press, 1996.

Dunne, Timothy, Mark Roberts and Larry Samuelson, "The Growth and Failure of U.S. Manufacturing Plants" *The Quarterly Journal of Economics*, November 1989.

Holmes, Thomas, and James Schmitz, "On the Turnover of Business Owners and Business Managers," *Journal of Political Economy*, October 1995.

Jarmin, Ron, and Javier Miranda, "The Longitudinal Business Database", 2003, CES Working Paper.

Nucci, Alfred and Richard Boden. 2003. "Demography of Nonemployer Businesses -- Preliminary Evidence from the United States." Paper Presented at the Comparative Analysis of Enterprise (micro) Data Conference, London, England, September 15-16, 2003

Table 2.1 (a) Employer Business Register

Variable	Source Type	Source Details	Line Number
Name and Address	Reported	Business Reported physical address in Company Organization Survey, Annual Surveys, or Economic Censuses in Census years	
	Administrative	Beginning in 1998, physical address from form ss-4 for births; before 1998, mailing address from form ss-4	4a,4b: 5a,5b
		Beginning in 1998, physical address from form 941 for all businesses; before 1998, mailing address from form 941 for all businesses	Top of Form
		Beginning in 1998, physical address from IRS income tax form for all businesses; before 1998, mailing address from IRS income tax form for all businesses	Top of Form
Industry	Reported	Business Reported industry code in Company Organization Survey, Annual Surveys, or Economic Censuses in Census years	
		Derived from the 1992 Economic Census--respondent reported classification	
		Derived from a current survey (County Business Patterns (CBP), Company Organization Survey (COS)/Annual Survey of Manufactured (ASM), Current Industrial Report (CIR), Business Sample Revision (BSR) CBP Safeguard Review or intercensal refiles	
	Administrative	Derived from the Bureau of Labor Statistics	
		Derived from the Social Security Administration	
		Derived from the Internal Revenue Service (PBA) code (Business Master File (BMF)	
Employment	Reported*	Business Reported employment in Company Organization Survey, Annual Surveys, or Economic Censuses in Census years	
	Administrative	Obtained from Form 941	Line 1
		IRS FICA wages inserted	
		IRS Total Compensation inserted	
		Imputed	
Payroll	Reported*	Business Reported employment in Company Organization Survey, Annual Surveys, or Economic Censuses in Census years	
	Administrative	Obtained from Form 941	Line 2
		IRS FICA wages inserted	
		IRS Total Compensation inserted	
		Imputed	
Revenue	Reported* Imp from EIN-level	Business Reported employment in Company Organization Survey, Annual Surveys, or Economic Censuses in Census years	
		1120 - Gross receipts or sales less returns and allowances	Line 1c
		1120-A - Gross receipts or sales less returns and allowances	Line 1c
		1120F - Gross receipts or sales less returns and allowances	Section II, line 1a
		1120L - Gross Income	Line 9
		1120-PC - Gross Income	Sch A: line 14
		1120-RIC - Total Income	Line 8
		1120S - Gross receipts or sales less returns and allowances	Line 1c
		1065 - Gross receipts or sales less returns and allowances	Line 1c
		990 - Total Revenue	Line 12
		990-C - Gross receipts or sales less returns and allowances	Line 1c
		990EZ - Total Revenue	Line 9
		990-PF - Total Revenue	Line 12
		1040C - Gross receipts or sales less returns and allowances	Line 3

* = Multi-Unit Establishments Only

Table 2.1(b) NonEmployer Business Register - SSN Records

Variable	Source	Source Details
Name	Sole Props Tax form 1040 C Line C	Profit or Loss from Business Name of Proprietor
Mailing Address	Line E	Line E
Legal Form of Organization	implied by filing of form 1040 schedule C less returns and allowances	
Industry Code	Line B	Principal Business or Profession, including product or service
Revenue	Line 3: Gross receipts or sales	Gross receipts or sales less returns

NOTE: All data derived from tax form 1040 Schedule C.

NOTE: Legal form of organization implied by type of form submitted:

NOTE: Industry codes on Census Nonemployer database are IRS Principal Business Activity codes obtained from the sources noted above and then converted to Census Tabulated Kind of Business. If PBA not available on tax form, historic TKB is used when available.

Table 2.1 (c) NonEmployer Business Register - Partnerships and Corporations

Form Number	Form Description	Industry Code Line Number	Revenue Line Number
Form 1065	U.S. Partnership Return of Income	A (Principal Business Activity	1c: Gross receipts or sales less returns and allowances
Form 1120	U.S. Corporation Income Tax Return	Schedule K, 2a: Business activity code no.	1c: Gross receipts or sales less returns and allowances
Form 1120A	U.S. Corporation Short Form Tax Return	Part 2, 1a: Business Activity Code Number	1c: Gross receipts or sales less returns and allowances
Form 1120 S	U.S. Income Tax Return for an S Corporation	B. Business Code Number	1c: Gross receipts or sales less returns and allowances
Form 1120 F	U.S. Income Tax Return of a Foreign Corporation	F1: Business activity code number	Section II, 1c: Gross receipts or sales less returns and allowances
Form 1120 PC	U.S. Casualty and Property Insurance Company income Tax Return	Schedule I, Line 2	Schedule A: line 14 (gross income)
Form 1120 L	U.S. Life Insurance Company Income Tax Return	Schedule M, 2a-c (Kind of Company, principal business)	9 Gross Income
Form 1120 RIC	U.S. Income Tax Return for Regulated Investment Companies	Inferred by Form Type	8: Total Income
Form 1120 REIT	U.S. Income Tax Return for Real Estate Investment Trusts	Inferred by Form Type	8: Total Income

NOTE: All data for these businesses derived from IRS income tax return. Filed by business.

NOTE: Name and Mailing Address taken from top of forms.

NOTE: Industry codes on Census Nonemployer database are IRS Principal Business Activity codes obtained from the sources noted above and then converted to Census Tabulated Kind of Business. If PBA not available on tax form, historic TKB is used when available.

NOTE: Legal form of organization implied by type of form submitted: 1065 filers are partnerships, all others are corporations.

Table 2.2: Summary Statistics for the Employer and Nonemployer Business Universes, 2000

Nonemployer Business Universe

	All Industries	40 Selected Industries	
	Number	Number	Percent
EIN Units	2.15	0.54	25%
SSN Units	13.38	6.84	51%
All	15.54	7.38	48%
	Revenue	Revenues	Percent
EIN Units	251.74	55.58	22%
SSN Units	459.53	199.87	43%
All	711.26	255.45	36%

Shares of Aggregate Business Revenue

Nonemployer Businesses		*Employer Businesses*	
EIN Units	SSN Units	Single-Unit	Multi-Unit
1.43%	2.61%	34.77%	61.19%

Shares of Business Revenue, 40 Selected Industries

Nonemployer Businesses		*Employer Businesses*	
EIN Units	SSN Units	Single-Unit	Multi-Unit
3.09%	11.12%	48.83%	36.96%

Employer Business Universe

	All Industries	40 Selected Industries	
	Number	Number	Percent
Single-Unit	5.26	1.90	36%
Multi-Unit	0.18	0.06	31%
All	5.44	1.96	36%
	Revenue	Revenue	Percent
Single-Unit	6,113.43	877.92	14%
Multi-Unit	10,758.04	664.52	6%
All	16,871.47	1,542.44	9%
	Payroll	Payroll	Percent
Single-Unit	1,425.88	293.76	21%
Multi-Unit	2,347.12	235.18	10%
All	3,773.00	528.94	14%
	Employment	Employment	Percent
Single-Unit	48.86	10.98	22%
Multi-Unit	64.79	7.77	12%
All	113.66	18.75	16%

Number of Units and Employment in Millions; Revenue and Payroll in Billions.

Table 2.3 Selected Industries in the Integrated Business Universe, 2000

	Employer Firms				NonEmployer *	
Label	Firms	Payroll	Employment	Revenue	Nonemployer Entities*	Revenue
Animal production support activities	1.5	156	7	664	37.7	1,462
Painting & wall covering contractors	36.4	6,274	226	17,469	213.5	7,443
Carpentry & floor contractors	52.3	9,637	383	35,398	389.6	16,722
Roofing, siding, & sheet metal contractors	27.6	8,143	264	27,315	86.5	5,047
Concrete contractors	24.5	5,092	178	18,185	42.6	2,567
Printing & related support activities	23.2	14,226	420	41,613	26.7	1,486
Ship & boat building	1.1	3,788	105	16,896	0.4	47
Gasoline stations	52.2	12,282	837	187,841	9.4	1,682
Book, periodical & music stores	9.4	2,713	222	12,577	28.9	1,008
Florists	20.7	1,645	123	6,417	22.7	869
Taxi & limousine service	5.1	1,206	65	3,451	117.6	3,419
Couriers	2.3	15,654	550	18,610	1.1	111
Software publishers	6.7	23,009	249	49,988	1.7	120
Agencies & other insurance related activities	101.7	30,448	745	90,461	308.2	14,849
Offices of real estate agents & brokers	50.0	11,504	297	46,826	476.6	22,952
Activities related to real estate	39.5	15,052	484	38,058	356.3	18,274
Consumer goods rental	13.7	3,483	236	11,391	16.0	768
Legal services	147.7	58,514	1,055	149,400	206.3	11,626
Accounting, tax prep, bookkeep, payroll service	76.9	21,273	765	42,205	294.5	6,059
Computer systems design & related services	74.9	76,674	1,194	141,900	249.4	9,688
Management, sci & tech consulting services	76.5	32,718	729	67,277	355.2	16,796
Travel arrangement & reservation services	21.3	8,713	292	45,546	31.8	1,776
Services to buildings & dwellings	115.2	24,903	1,407	59,490	538.9	11,294
Offices of physicians	120.5	46,346	1,043	102,651	149.6	11,664
Offices of dentists	35.0	7,268	273	19,865	29.4	1,907
Offices of other health practitioners	74.7	8,814	346	31,356	235.2	9,053
Individual & family services	21.1	9,411	478	11,919	65.3	1,096
Child day care services	39.2	9,747	691	14,125	516.6	6,263
Agents, managers for artists & other public fig	1.8	502	10	1,549	25.0	867
Independent artists, writers & performers	8.9	2,272	38	4,461	465.1	9,631
Rooming & boarding houses	1.5	182	13	717	9.7	281
Full-service restaurants	91.9	20,908	1,863	47,763	29.0	3,308
Limited-service eating places	94.4	17,164	1,743	53,707	36.8	3,086
Special food services	11.2	8,314	563	17,466	68.5	2,014
Drinking places (alcoholic beverages)	44.3	3,753	360	12,468	21.3	1,466
Automotive repair & maintenance	94.8	12,533	583	46,945	251.2	11,570
Personal & household goods R&M	22.4	2,337	98	8,685	247.9	6,147
Personal care services	59.8	4,109	303	9,770	552.4	11,776
Drycleaning & laundry services	30.8	6,613	373	17,731	33.0	1,650
Other personal services	16.0	3,688	173	12,288	835.8	17,609
Total Economy	5,443.40	3,773,003	113,658	16,871,471	15,536.07	711,264

Firms and employment in thousands. Payroll and Revenue in millions.

Table 2.4 Revenue Shares for Employers and Nonemployers, 40 Selected Industries, 2000

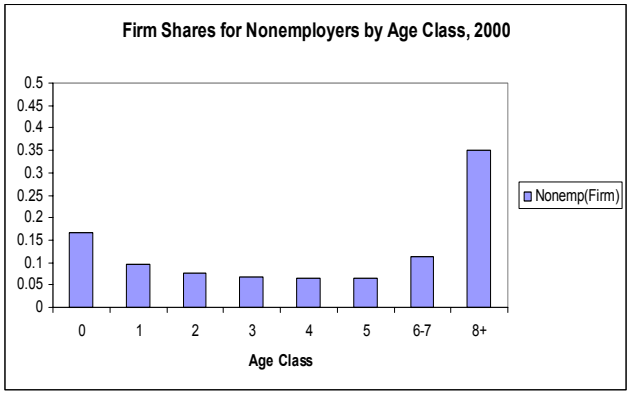
NAICS	Label	Industry Share of Aggregate Business Revenue	Employers' Share of Industry Revenues	Nonemployers' Share of Industry Revenues
1152	Animal production support activities	0.01%	31%	69%
2352	Painting & wall covering contractors	0.14%	70%	30%
2355	Carpentry & floor contractors	0.30%	68%	32%
2356	Roofing, siding, & sheet metal contractors	0.18%	84%	16%
2357	Concrete contractors	0.12%	88%	12%
3231	Printing & related support activities	0.25%	97%	3%
3366	Ship & boat building	0.10%	100%	0%
4471	Gasoline stations	1.08%	99%	1%
4512	Book, periodical & music stores	0.08%	93%	7%
4531	Florists	0.04%	88%	12%
4853	Taxi & limousine service	0.04%	50%	50%
4921	Couriers	0.11%	99%	1%
5112	Software publishers	0.28%	100%	0%
5242	Agencies & other insurance related activities	0.60%	86%	14%
5312	Offices of real estate agents & brokers	0.40%	67%	33%
5313	Activities related to real estate	0.32%	68%	32%
5322	Consumer goods rental	0.07%	94%	6%
5411	Legal services	0.92%	93%	7%
5412	Accounting, tax prep, bookkeep, payroll service	0.27%	87%	13%
5415	Computer systems design & related services	0.86%	94%	6%
5416	Management, sci & tech consulting services	0.48%	80%	20%
5615	Travel arrangement & reservation services	0.27%	96%	4%
5617	Services to buildings & dwellings	0.40%	84%	16%
6211	Offices of physicians	0.65%	90%	10%
6212	Offices of dentists	0.12%	91%	9%
6213	Offices of other health practitioners	0.23%	78%	22%
6241	Individual & family services	0.07%	92%	8%
6244	Child day care services	0.12%	69%	31%
7114	Agents, managers for artists & other public fig	0.01%	64%	36%
7115	Independent artists, writers & performers	0.08%	32%	68%
7213	Rooming & boarding houses	0.01%	72%	28%
7221	Full-service restaurants	0.29%	94%	6%
7222	Limited-service eating places	0.32%	95%	5%
7223	Special food services	0.11%	90%	10%
7224	Drinking places (alcoholic beverages)	0.08%	89%	11%
8111	Automotive repair & maintenance	0.33%	80%	20%
8114	Personal & household goods R&M	0.08%	59%	41%
8121	Personal care services	0.12%	45%	55%
8123	Drycleaning & laundry services	0.11%	91%	9%
8129	Other personal services	0.17%	41%	59%

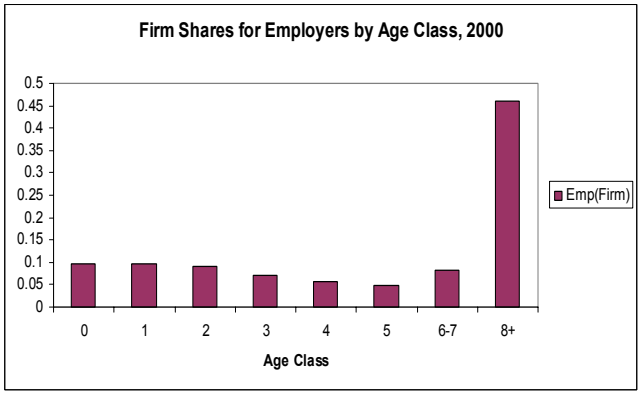
Table 3.1 Summary Statistics on Backward Linkages from Employers to Nonemployers

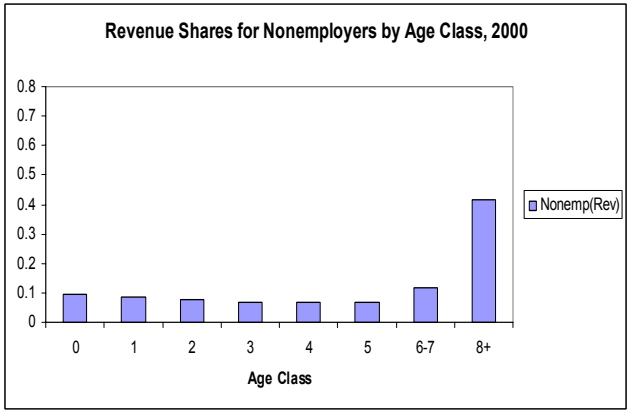
Firm Size	# of Firms	% of Firms	% of total 2000 LBD Payroll	% of last year NE Receipts	# of firms with links	% of total links	% of firms with links
a)1-4	1416292	61.82%	9.81%	46.73%	200252	75.36%	14.14%
b)5-9	432027	18.86%	9.02%	13.09%	34590	13.02%	8.01%
c)10-19	242636	10.59%	10.58%	12.38%	16656	6.27%	6.86%
d)20-49	139368	6.08%	13.43%	8.62%	9229	3.47%	6.62%
e) 50-99	36886	1.61%	8.56%	6.18%	2526	0.95%	6.85%
f)100-249	16426	0.72%	8.83%	4.48%	1362	0.51%	8.29%
g)250-499	4332	0.19%	7.16%	2.32%	489	0.18%	11.29%
h) 500+	3103	0.14%	32.62%	6.20%	634	0.24%	20.43%
Total	2291070	100.00%	100.00%	100.00%	265738		11.60%

Firm Age	# of Firms	% of Firms	% of total 2000 LBD Payroll	% of last year NE Receipts	# of firms with links	% of total links	% of Fr with links
0	213292	9.31%	2.28%	11.01%	35082	13.20%	16.45%
1	190690	8.32%	3.27%	9.08%	31314	11.78%	16.42%
2	170091	7.42%	3.40%	7.95%	27197	10.23%	15.99%
3	153400	6.70%	3.44%	6.46%	25244	9.50%	16.46%
4	134315	5.86%	3.40%	5.49%	20675	7.78%	15.39%
5	117723	5.14%	3.10%	4.48%	20615	7.76%	17.51%
6 – 7	199523	8.71%	5.61%	10.81%	24823	9.34%	12.44%
8+	1112036	48.54%	75.50%	44.74%	80788	30.40%	7.26%
Total	2291070	100.00%	100.00%	100.00%	265738		11.60%

Figure 2.1


Firm Shares for Nonemployers by Age Class, 2000
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6-7 8+
Age Class
Nonemp(Firm)


Firm Shares for Employers by Age Class, 2000
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6-7 8+
Age Class
Emp(Firm)


Revenue Shares for Nonemployers by Age Class, 2000
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6-7 8+
Age Class
Nonemp(Rev)

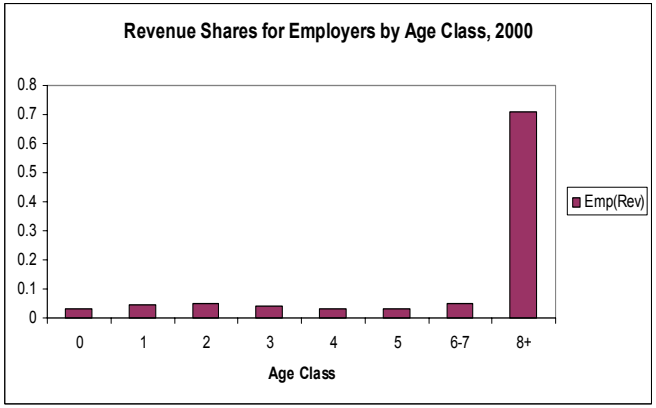

Revenue Shares for Employers by Age Class, 2000
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6-7 8+
Age Class
Emp(Rev)

Figure 2.2

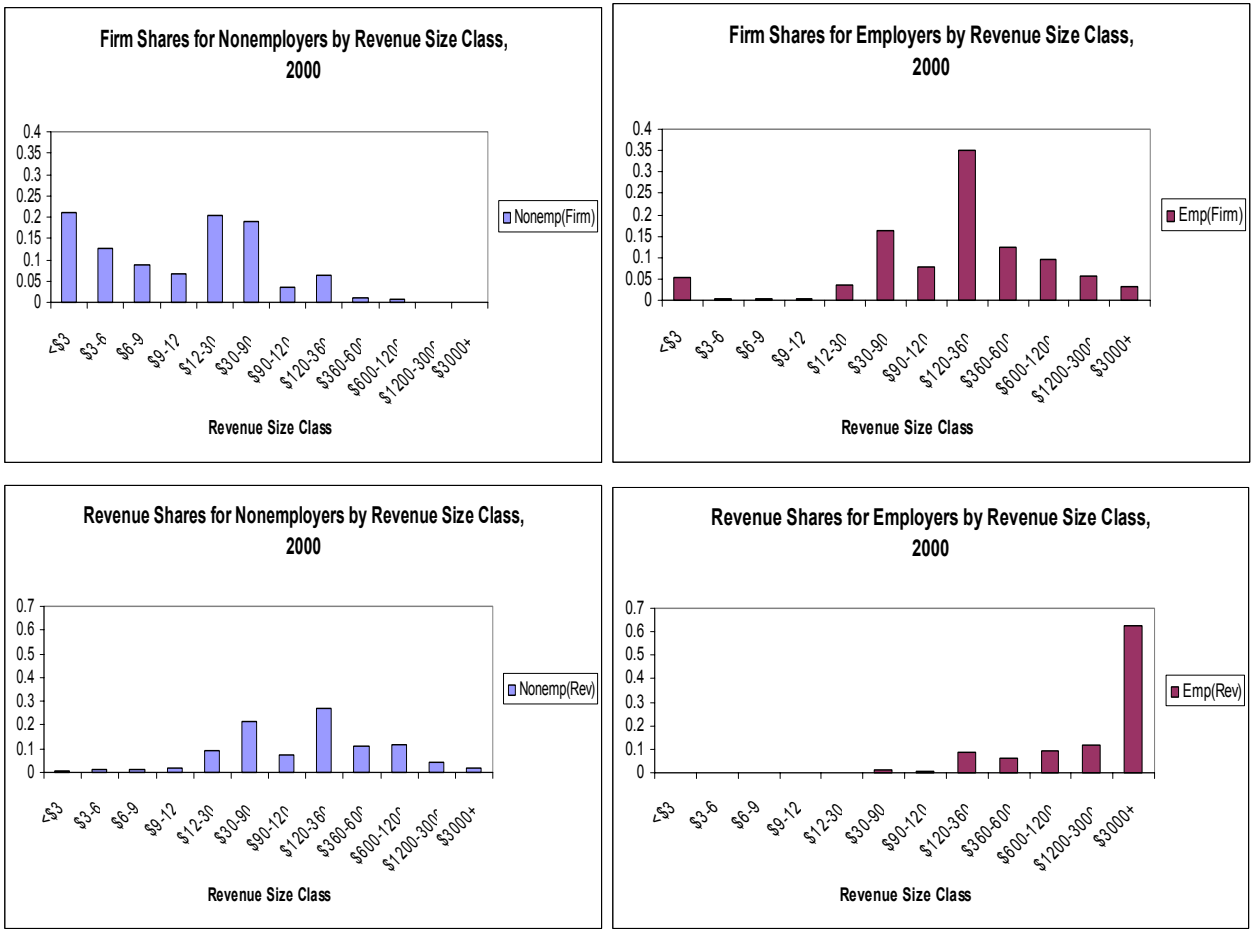
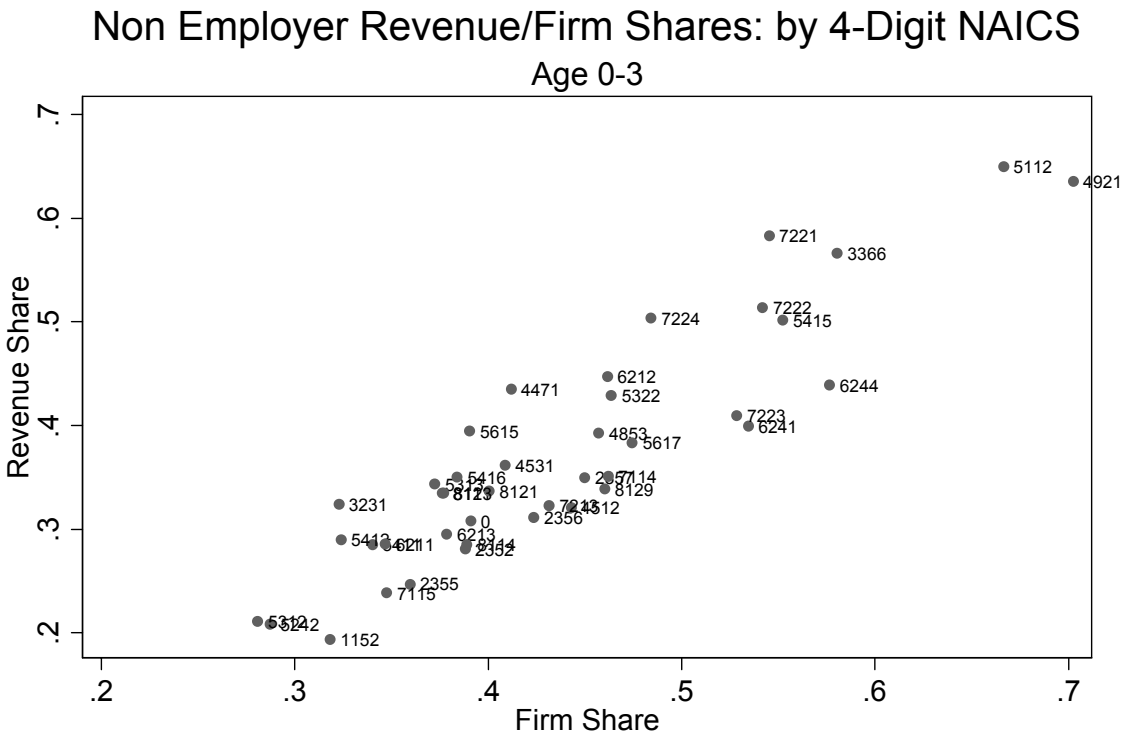
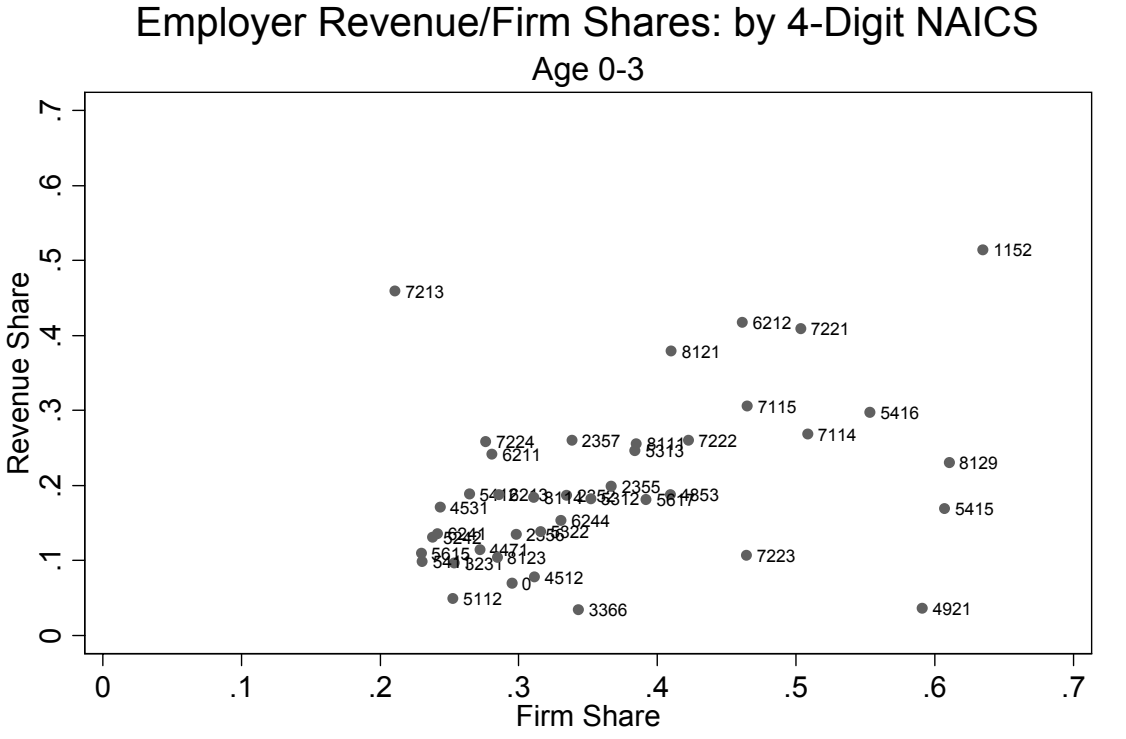

Firm Shares for Nonemployers by Revenue Size Class, 2000
Nonemp(Firm)
Revenue Size Class
Firm Shares for Employers by Revenue Size Class, 2000
Emp(Firm)
Revenue Size Class
Revenue Shares for Nonemployers by Revenue Size Class, 2000
Nonemp(Rev)
Revenue Size Class
Revenue Shares for Employers by Revenue Size Class, 2000
Emp(Rev)
Revenue Size Class

Figure 2.3

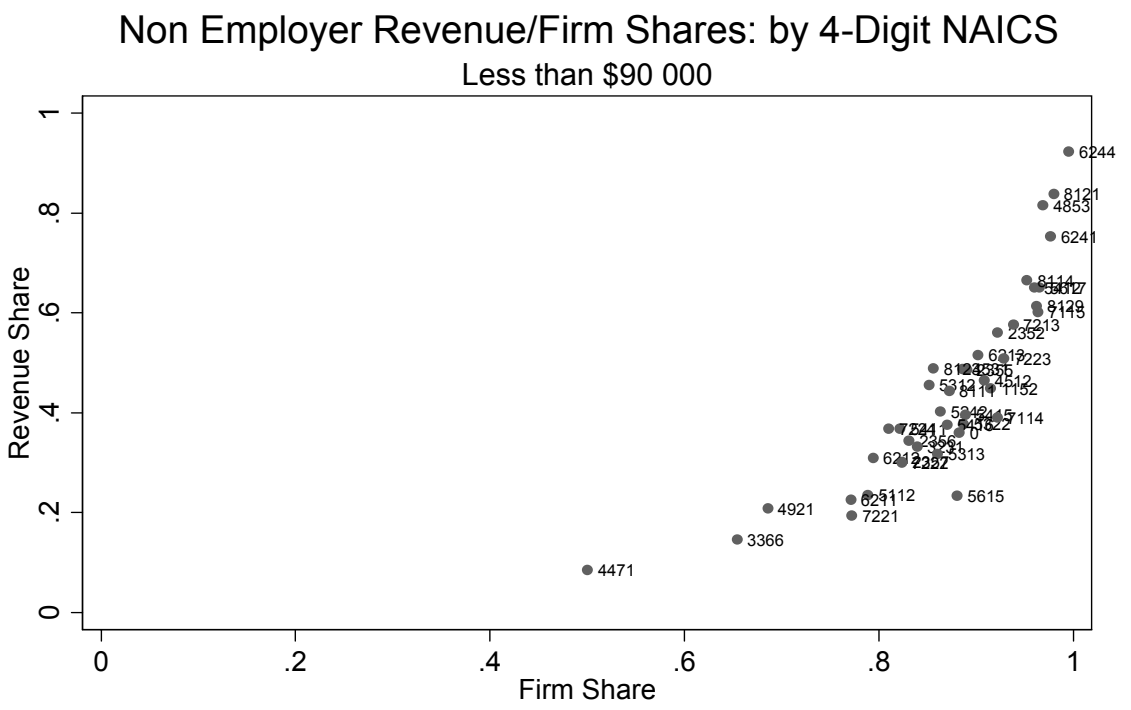


Source: Own Calculations from ILBD

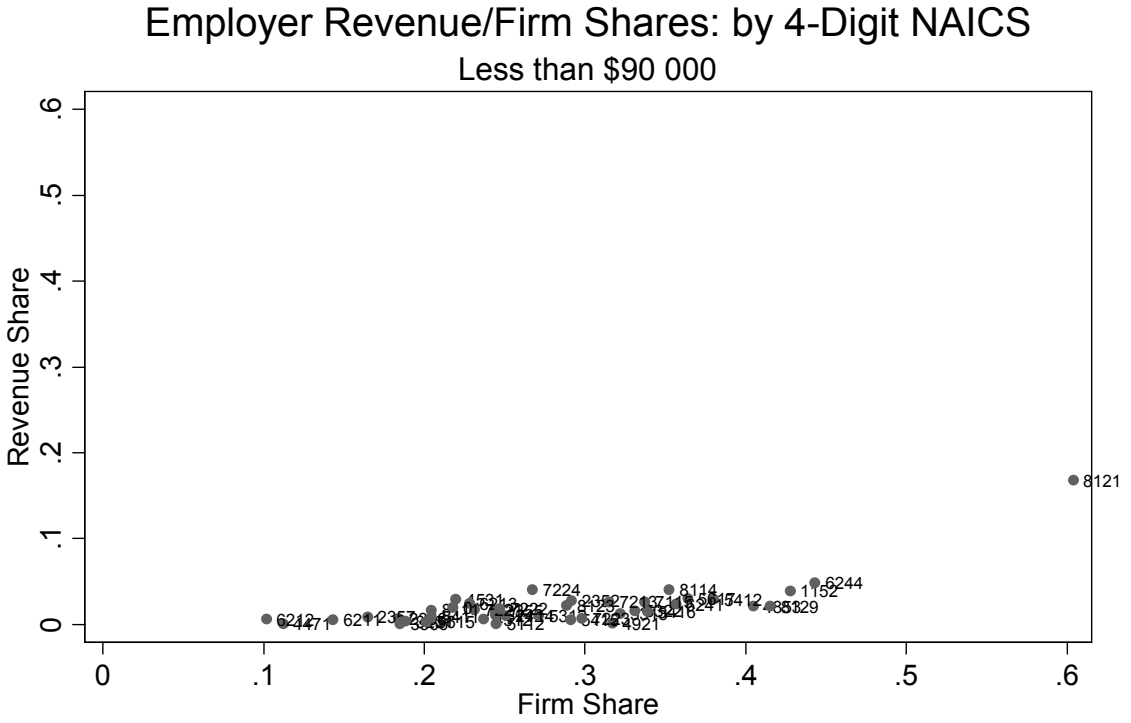


Source: Own Calculations from ILBD

Figure 2.4



Source: Own Calculations from ILBD



Source: Own Calculations from ILBD

Figure 3.1

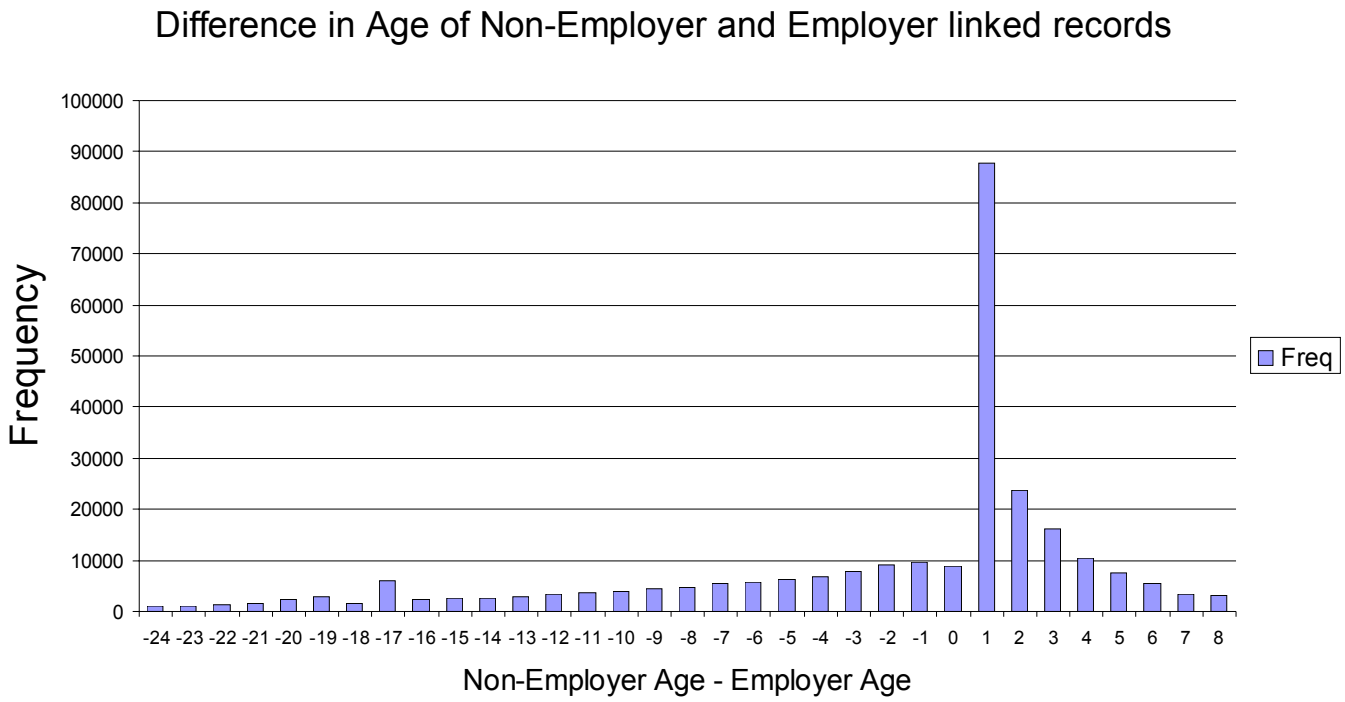
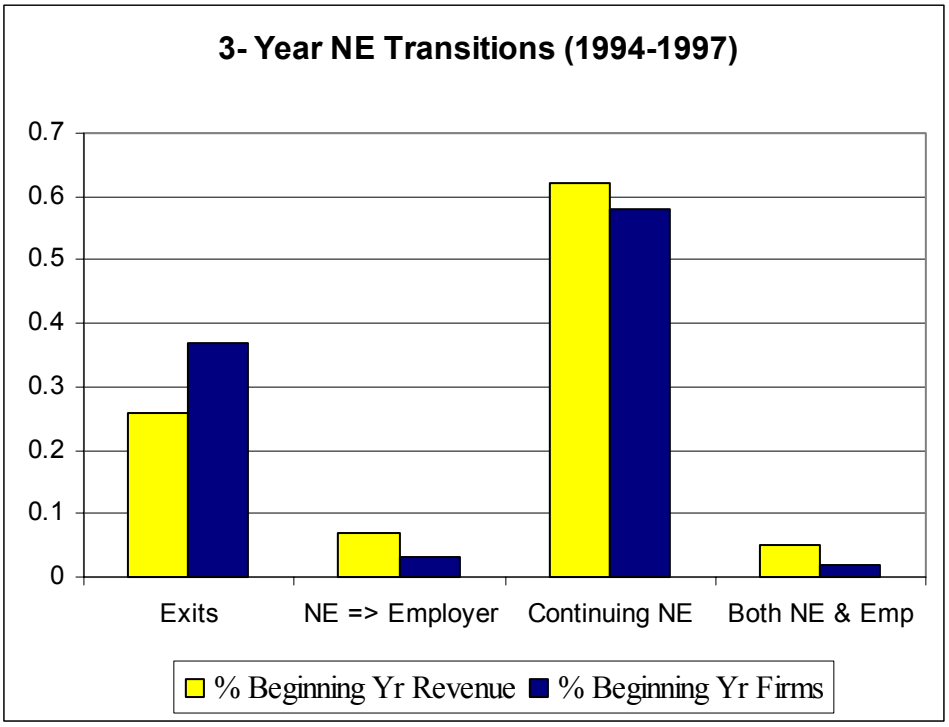
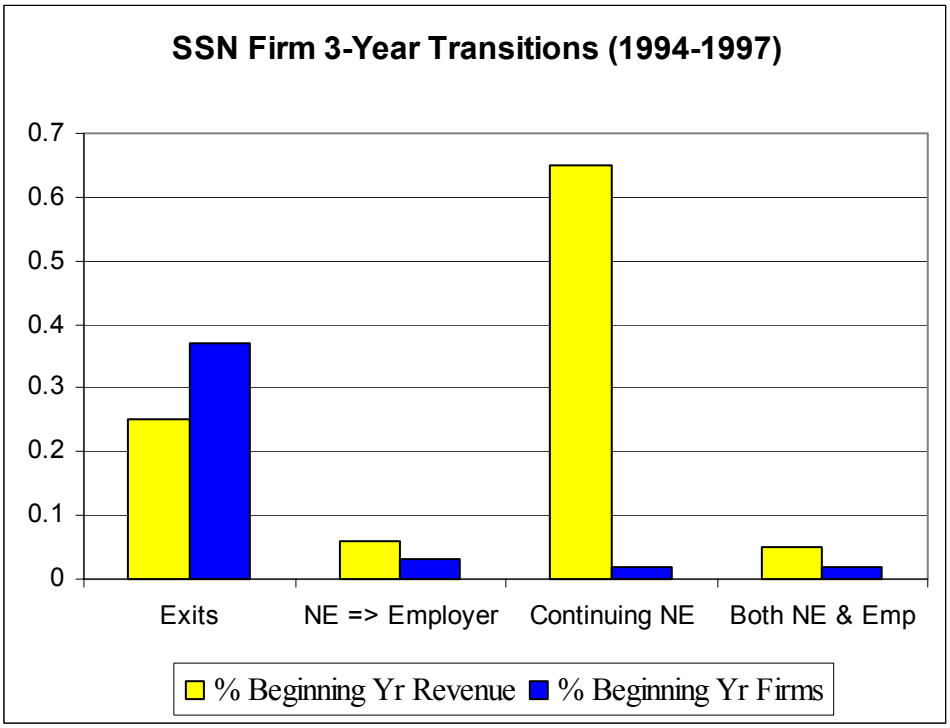

Difference in Age of Non-Employer and Employer linked records
Frequency
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
-24 -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8
Non-Employer Age - Employer Age
Freq

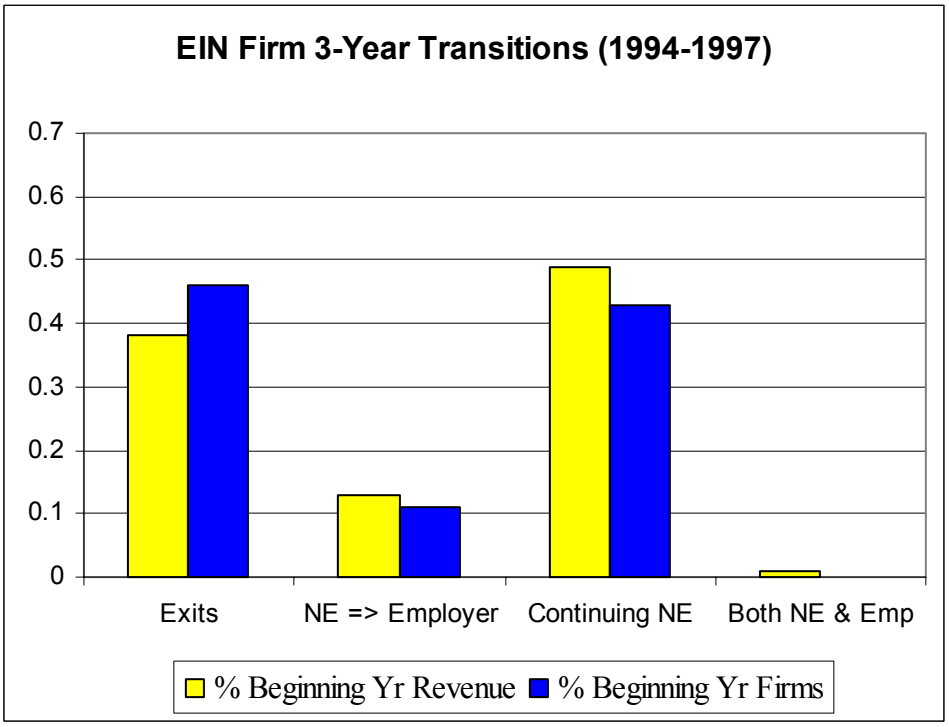
Figure 4.1



Note: NE=Nonemployer, Emp=Employer

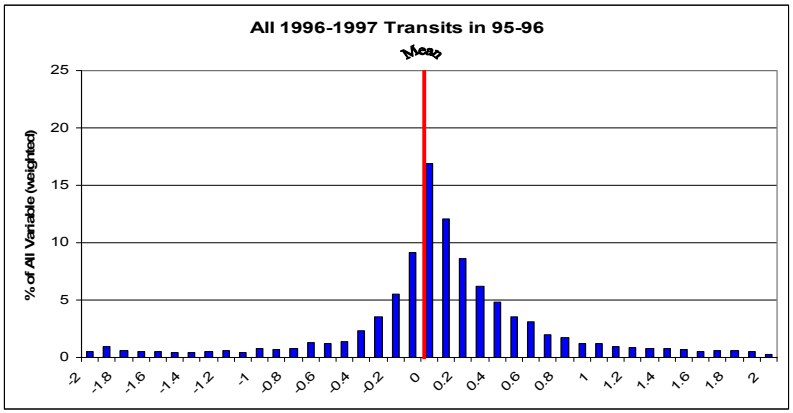
Figure 4.2

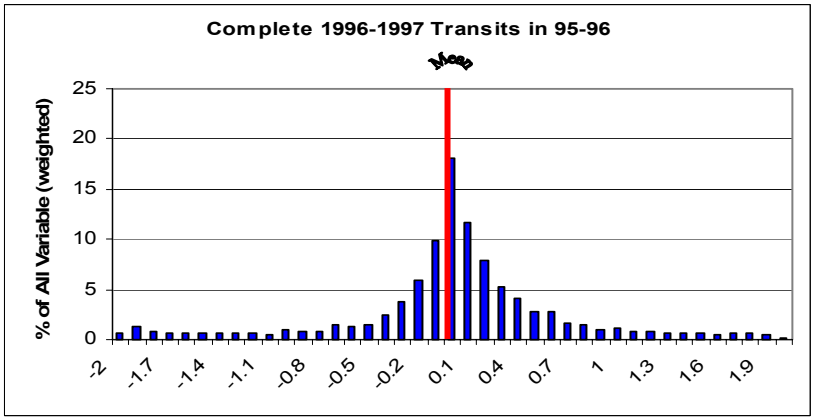


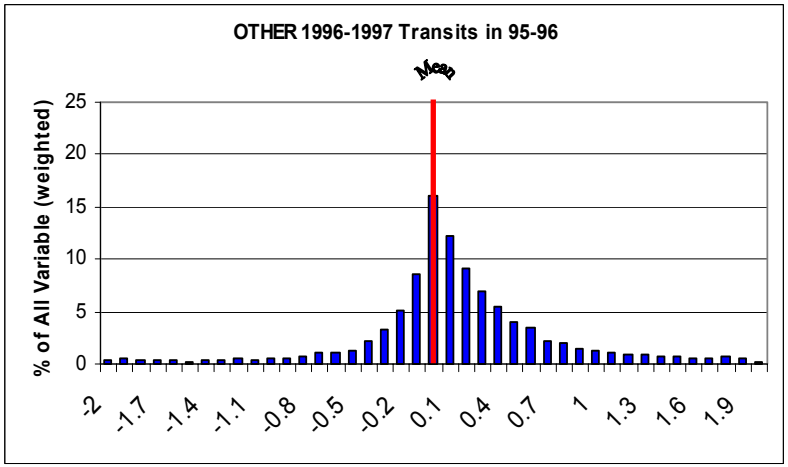


Note: NE=Nonemployer, Emp=Employer

Figure 4.3


All 1996-1997 Transits in 95-96
Mean
% of All Variable (weighted)
25
20
15
10
5
0
-2
-1.8
-1.6
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2


Complete 1996-1997 Transits in 95-96
Mean
% of All Variable (weighted)
25
20
15
10
5
0
-2
-1.7
-1.4
-1.1
-0.8
-0.5
-0.2
0.1
0.4
0.7
1
1.3
1.6
1.9


OTHER 1996-1997 Transits in 95-96
Mean
% of All Variable (weighted)
25
20
15
10
5
0
-2
-1.7
-1.4
-1.1
-0.8
-0.5
-0.2
0.1
0.4
0.7
1
1.3
1.6
1.9

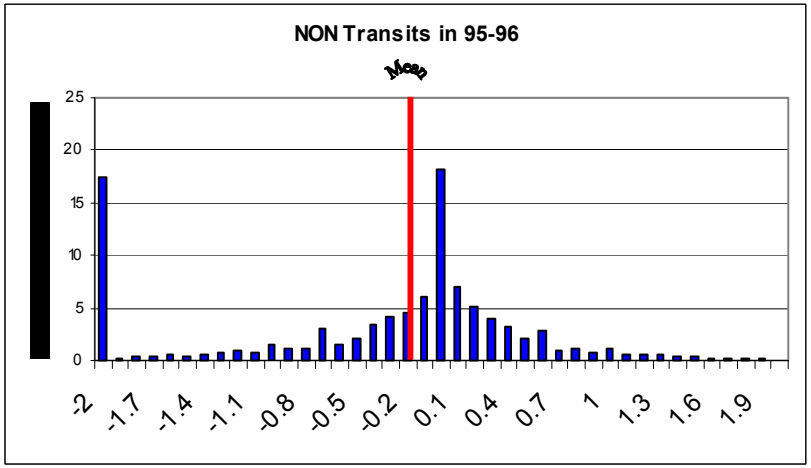
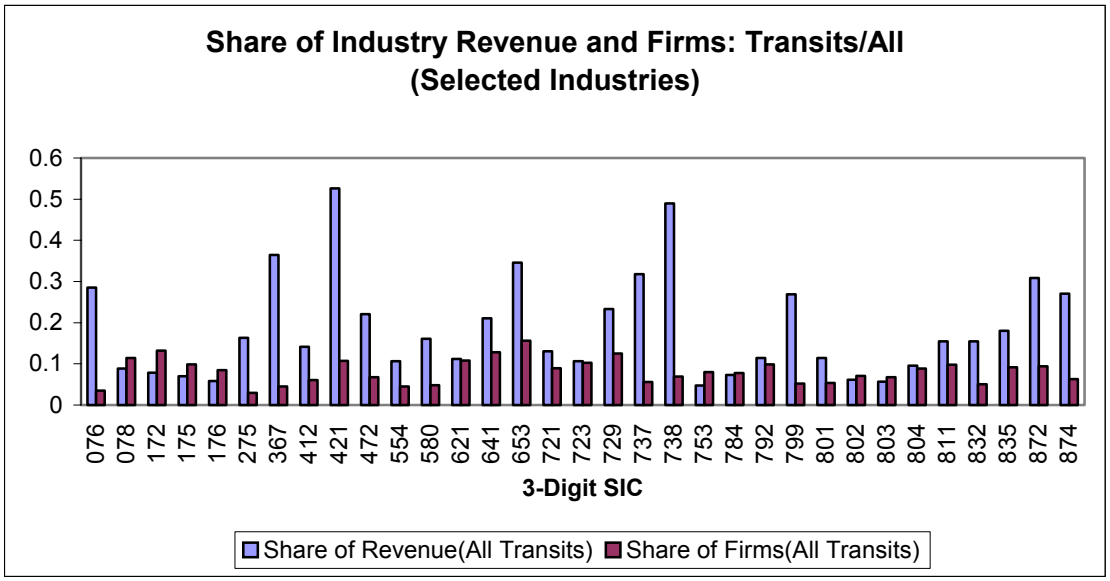

NON Transits in 95-96
Mean
25
20
15
10
5
0
-2
-1.7
-1.4
-1.1
-0.8
-0.5
-0.2
0.1
0.4
0.7
1
1.3
1.6
1.9

Figure 4.4


Share of Industry Revenue and Firms: Transits/All
(Selected Industries)
0.6
0.5
0.4
0.3
0.2
0.1
0
076 078 172 175 176 275 367 412 421 472 554 580 621 641 653 721 723 729 737 738 753 784 792 799 801 802 803 804 811 832 835 872 874
3-Digit SIC
Share of Revenue(All Transits)
Share of Firms(All Transits)

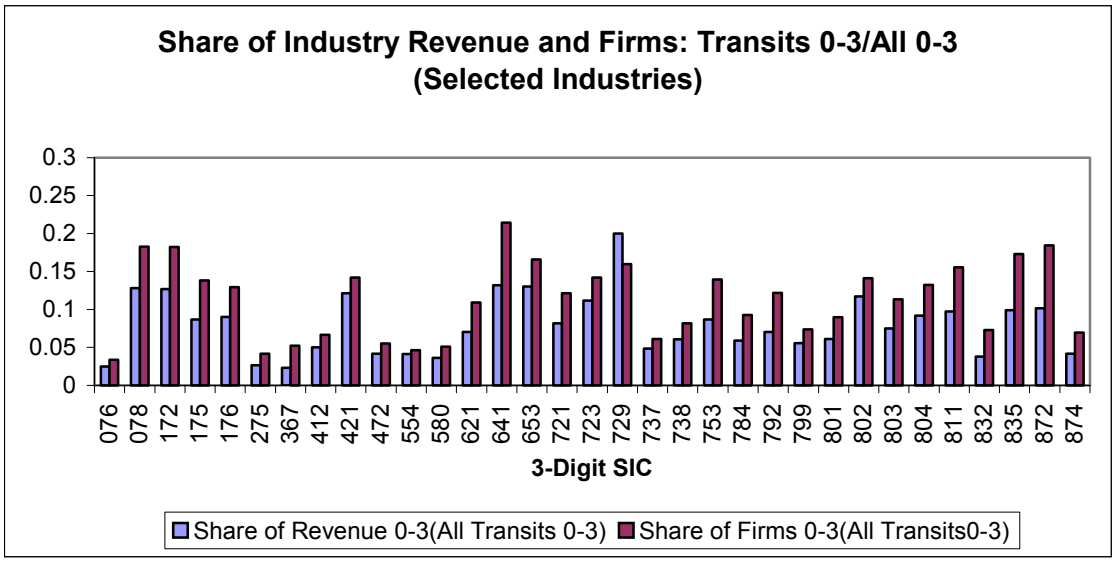
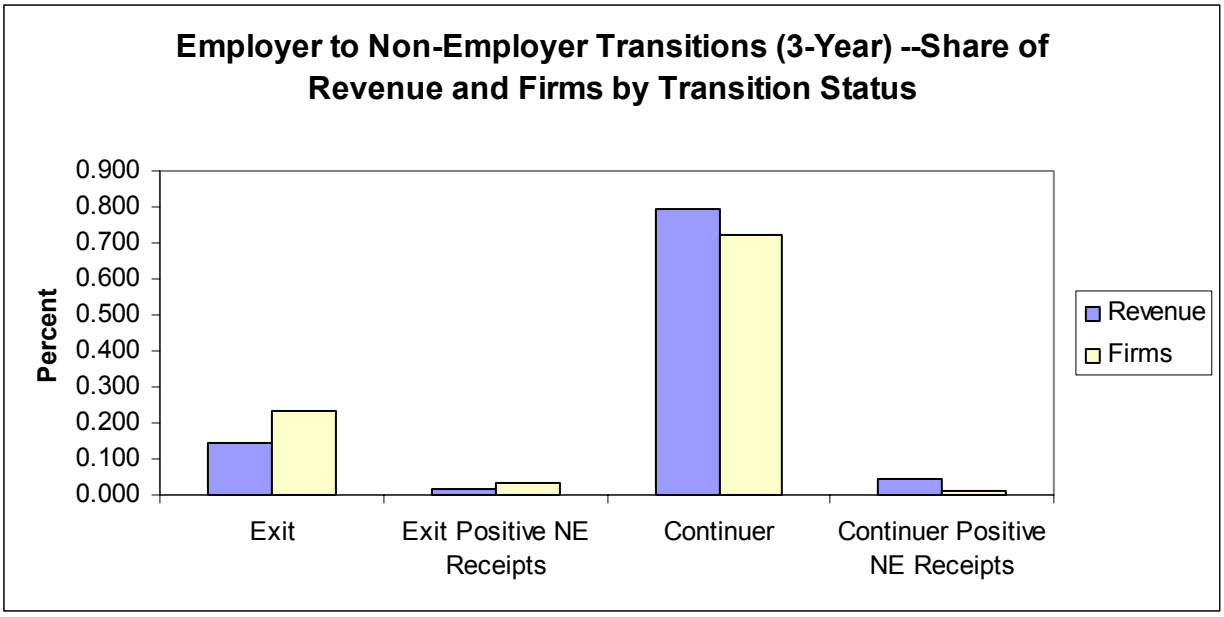
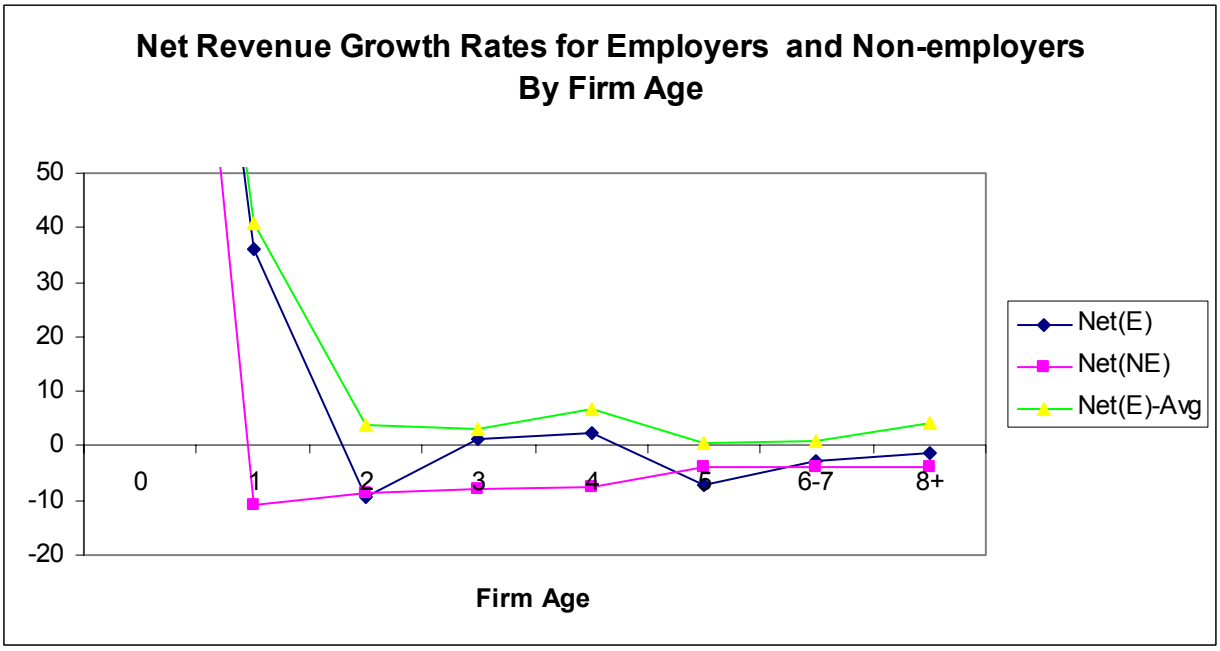

Share of Industry Revenue and Firms: Transits 0-3/All 0-3
(Selected Industries)
0.3
0.25
0.2
0.15
0.1
0.05
0
076 078 172 175 176 275 367 412 421 472 554 580 621 641 653 721 723 729 737 738 753 784 792 799 801 802 803 804 811 832 835 872 874
3-Digit SIC
Share of Revenue 0-3(All Transits 0-3)
Share of Firms 0-3(All Transits0-3)

Figure 4.5



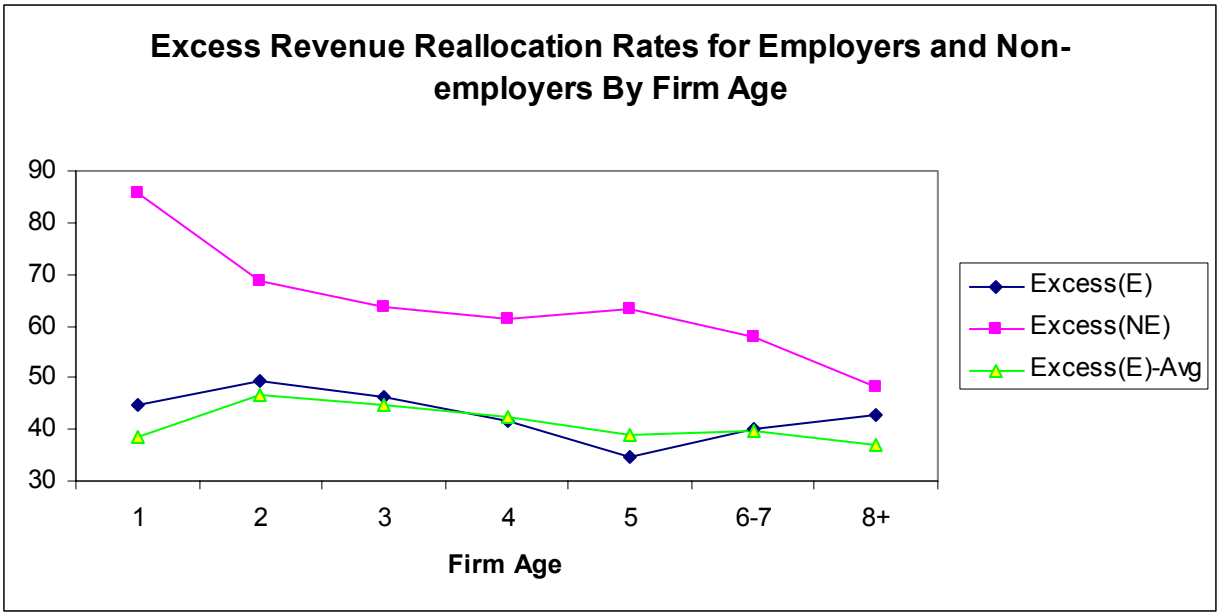
Note: NE=Nonemployer

Figure 5.1



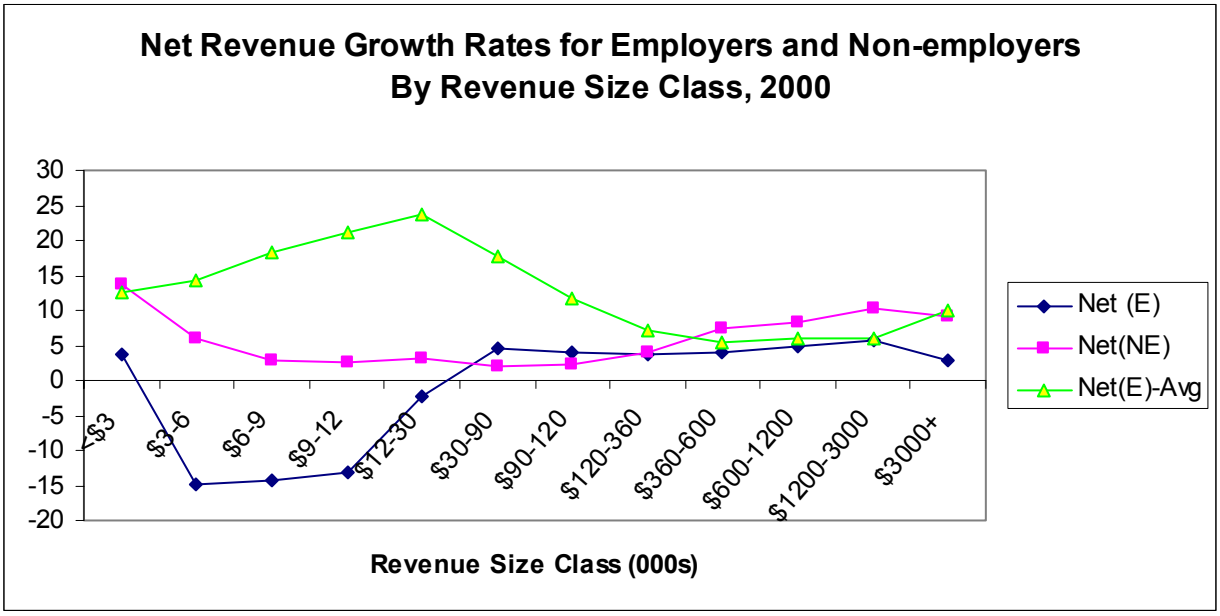
Note: NE=Nonemployer, E=Employer. Net(E) and Net(NE) represent net revenue growth rate for age group for 1999-2000. Net(E)-Avg is the average net revenue growth rate for the age group over the years 1994-2000.

Figure 5.2



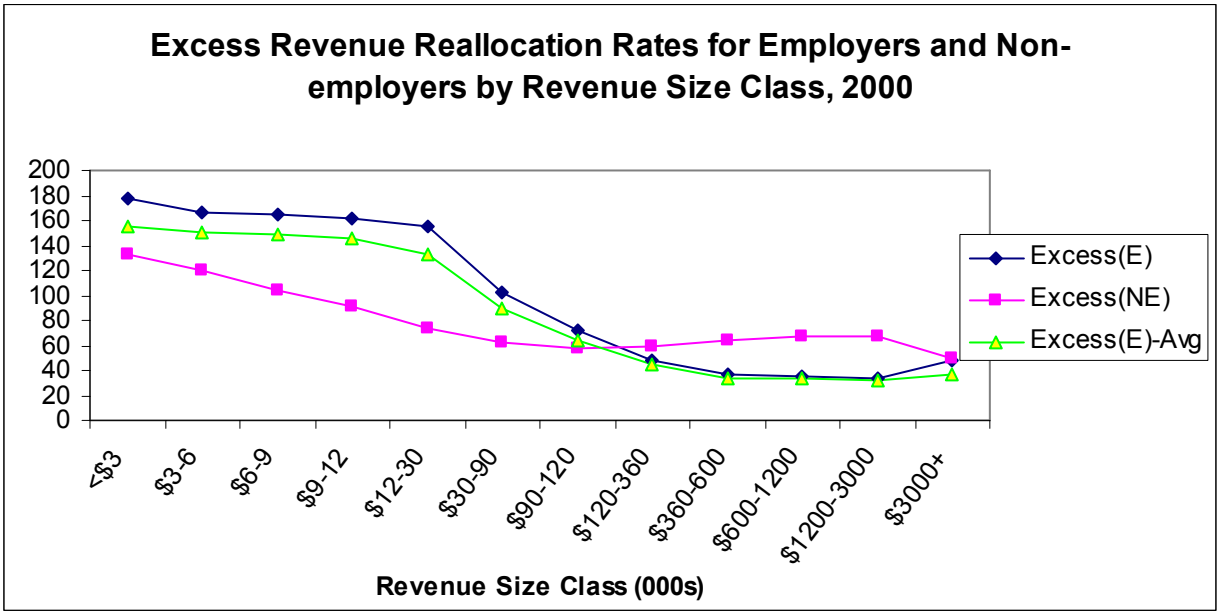
Note: NE=Nonemployer, E=Employer. Excess(E) and Excess(NE) are the excess revenue reallocation rates for the age group for 1999-2000. Excess(E)-Avg is the average excess revenue reallocation rate for the age group over the years 1994-2000.

Figure 5.3



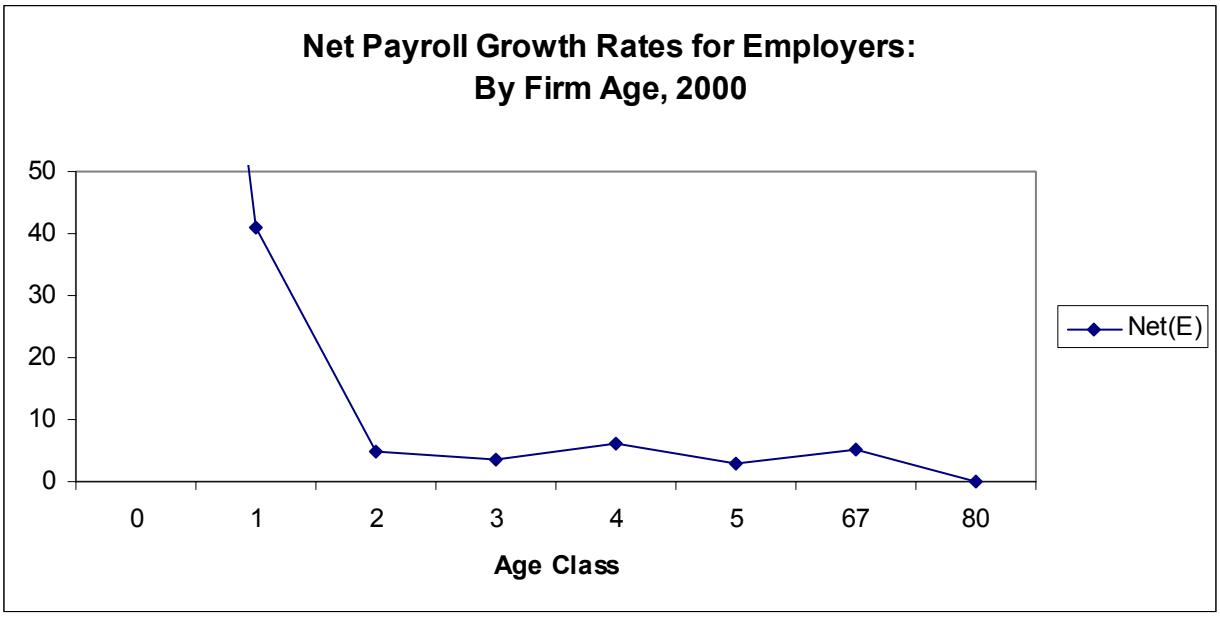
Note: NE=Nonemployer, E=Employer. Net(E) and Net(NE) represent net revenue growth rate for size group for 1999-2000. Net(E)-Avg is the Average Net Revenue Growth Rate for the size group over the years 1994-2000.

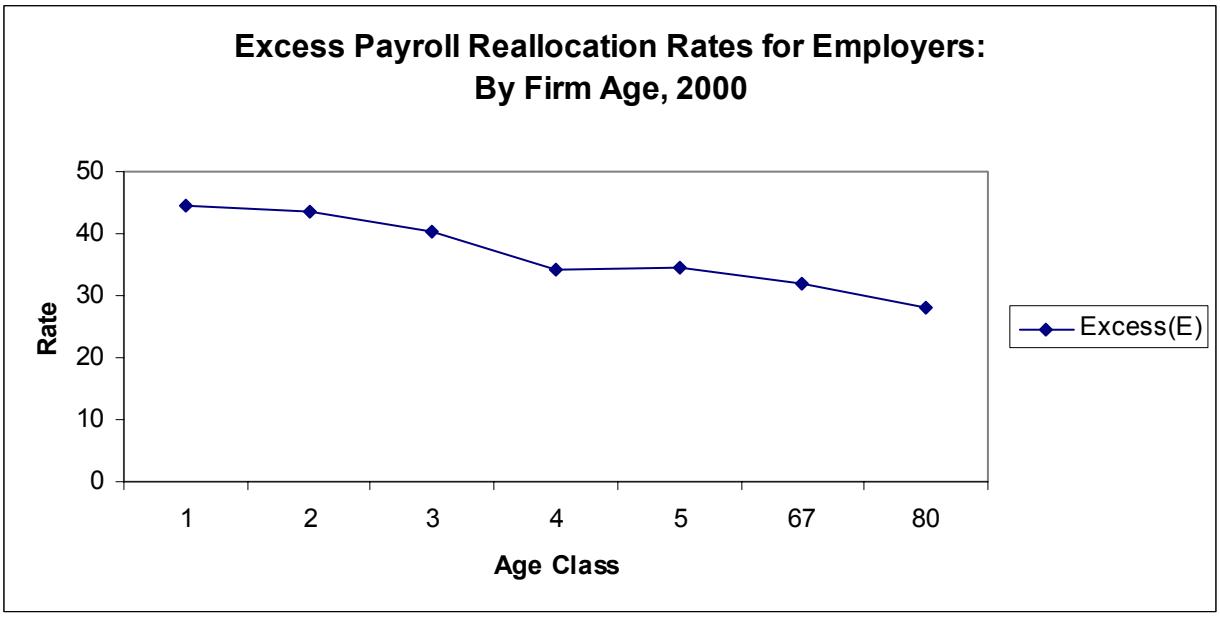
Figure 5.4



Note: NE=Nonemployer, E=Employer. Excess(E) and Excess(NE) are the excess revenue reallocation rates for the size group for 1999-2000. Excess(E)-Avg is the Average Excess Revenue Reallocation Rate for the size group over the years 1994-2000.

Figure 5.5


Net Payroll Growth Rates for Employers:
By Firm Age, 2000
50
40
30
20
10
0
0
1
2
3
4
5
67
80
Age Class
Net(E)


Excess Payroll Reallocation Rates for Employers:
By Firm Age, 2000
Rate
50
40
30
20
10
0
1
2
3
4
5
67
80
Age Class
Excess(E)