namics of young and small businesses. Major strengths of the ILBD include comprehensive industry and geographic coverage, longitudinal links for establishments and firms, linkability to the large number of business surveys housed at Census, and an integrated treatment of employer and nonemployer business. The ILBD makes it possible to examine the behavior over time of virtually all businesses in the U.S. economy, employers and nonemployers alike, with robust samples and even entire populations.

References


Comment

Thomas J. Holmes

Over the past two decades, much progress has been made in the analysis of business dynamics. For the most part, the business population studied in this literature is the population of employers as opposed to nonemployers. Statistical agencies, like the U.S. Census Bureau, place the vast bulk of their data collection energies on employers and so that is the vast bulk of data available to researchers. One particularly important employer data set

Thomas J. Holmes is the Curtis L. Carlson Professor of Economics at the University of Minnesota, a visiting scholar at the Federal Reserve Bank of Minneapolis, and a research associate of the National Bureau of Economic Research.
is the Longitudinal Research Database (LRD), which matches annual survey data of employers with the quinquennial economic census information. The analysis of the LRD, including work by the some of the authors of this chapter, has had a major impact on the profession.

Nonemployers are businesses operated by owner-managers (i.e., the subset of self-employed or partnership businesses with no employees). There exists a fairly large literature on self-employment. This literature is based mainly on surveys of individuals, such as the Current Population Survey (see for example, Evans and Leighton [1989]). But it also includes work based on federal income tax filings. Individuals who file a Schedule C (Proprietorship), a Partnership return, and a special kind of corporate return (S Corporation) when filing their personal income tax forms all are business owners of some kind. Some work uses the tax forms directly (Holtz-Eakin, Joulfaian, and Rosen 1994) or in conjunction with Census surveys, such as the Characteristics of Business Owner’s Survey that use the tax form population as the underlying business universe to sample from (Holmes and Schmitz 1995).

This research team is developing the Integrated Longitudinal Business Data Base (ILDB) and the chapter reports its initial findings. The project makes two data contributions. First, it links the tax-based information about nonemployer businesses over time. The chapter does not emphasize this first contribution much, but there is potentially significant research potential in being able to follow nonemployers over time, even ones who never become employers. Second, and this is what the chapter emphasizes, it links the nonemployer records with records in the employer files, such as the records in the LRD.

This is a very interesting project and the authors should be commended for engaging in this labor-intensive effort. Given all the important work that has been done on firm birth and growth with the employer population, it is natural to try to find out information about a firm before it even hires its first employee and gets in the employer data set. Basically, the chapter finds that about 16 percent of all young employer firms can be traced as having a backward link in the nonemployer file. So in thinking about the life cycle of the firm, they do find some evidence of firms starting as a nonemployer and then later adding employees and entering the employer data set. But interestingly, (and perhaps reassuringly for those who want to focus on employer data sets) this happens only a small percentage of the time. The overwhelming majority of employer firms do not have nonemployer antecedents. Apparently, firms that are eventually going to employ somebody start out with employees in the vast majority of cases.

The chapter shows that the links between the nonemployer and employer databases can form a complex web. As discussed previously, they find cases where a nonemployer in one year transits to employer status the following year, an easy case to understand. But they also find cases where
large employer firms that have existed for many years have concurrent links with nonemployer firms. This can happen, for example, if a social security number for an owner of a nonemployer business is linked to an employer business. The employer business will have social security numbers in its file because business owners need to report them when initially filing for an Employer Identification Number (EIN). There are a number of ways these kinds of links can appear. First, an employer might want to organize a nonemployer subsidiary for tax or regulatory purposes. Second, an original founder of an employer business might leave and later start a new nonemployer business. There are a myriad of possibilities. Down the line, it would be useful if progress can be made in differentiating between some of the various factors accounting for the links between the employer and nonemployer universes.

In conclusion, this research team has tackled a tough and messy project. It is messy because the underlying objects being studied are complex with complicated linkages. And there are more than the usual problems with linking data fields because of name changes, abbreviations, and so forth. While the problem is tough, I cannot think of a better team to take on the challenge. It is a first-rate team with extensive experience in this kind of data construction process. The team has put together an interesting data set that will be of use to researchers studying transitions from nonemployer status to employer status, as well as to researchers studying nonemployers for their own sake. While the project still seems in its early phases, the preliminary results are sensible. The results pass a “sniff test” that the matching exercise is working.

References