QUALITY OF U.S. BUSINESS ESTABLISHMENT FRAMES: DISCUSSION

Brenda G. Cox, Mathematica Policy Research, Inc.
600 Maryland Avenue, SW, Suite 550, Washington, DC 20024-2512

Creating and maintaining sampling frames for business surveys is difficult and expensive. Many problems relate to the unit being measured. A business is not a natural unit like a person or household that can be neatly or crisply defined. Businesses add employees and locations, lose employees and locations, and in some cases have no employees or locations. The rate of change in business organizations is high and they begin operations, change ownership, and go out of business at surprising rates.

In particular, small businesses are particularly difficult to maintain in a business register. The point at which they can be said to have started business can be imprecise, with new businesses being incubated out of homes with the only "employee" the owners who may not consider themselves to be employees. Capturing small businesses that die can be equally difficult. The end of a small business can be no more dramatic than turning off the lights and disconnecting the telephone.

On the other extreme, large businesses have complex ownership and management structures that make it difficult to define the characteristics of the business entity. Large enterprises buy independent companies, sell component companies, and merge operations with other organizations. What constitutes the "enterprise" can become murky, especially for the large, complexly organized entities.

These examples illustrate why building and maintaining a high quality business register resembles forging through a swamp riddled with quicksand. Each step must be carefully thought out and the end objectives must be kept in mind. The authors in this session discuss these three business registers which most Federal agencies use as sampling frames for business surveys:

- The Business Establishment List (BEL) maintained by the U.S. Bureau of Labor Statistics (BLS).

- The Standard Statistical Establishment List (SSEL) maintained by the U.S. Bureau of the Census.

- The Dun's Market Identifiers File (DMI) maintained by Dun & Bradstreet.

With some exceptions, use of the BEL and SSEL is limited to the maintenance agencies. Other agencies of the Federal government tend to use a private vendor such as Dun & Bradstreet to obtain their business frames.

The Business Establishment List

Searson and Farmer (1997) present an excellent description of the methodology underlying the Business Establishment List or BEL. The BEL is an establishment-level business register compiled from administrative reports submitted as a part of the Unemployment Insurance (UI) program. As such, it demonstrates the complexities and coordination problems associated with administrative databases as well as the advantages. The complexities are associated with the UI laws which vary across states and result in variations in terms of exactly what businesses (and employees) are covered. The advantages relate to having an efficient mechanism for identifying (1) businesses just starting operations, (2) changes in existing businesses, and (3) businesses that have ceased operations.

Searson and Farmer's discussion of coverage focused on percent of employees covered. I would have also liked to have heard about the BEL's coverage of businesses. The BEL covers 98 percent of non-farm employment, for instance. What fraction of non-farm employers does it cover? Business registers tend to have more difficulty capturing small businesses which constitute a large portion of the nation's total business but not nearly the same proportion of the nation's total employees. Because inclusion equates to UI taxes in this instance, small businesses may also have incentives to neglect to report their existence.

I find it curious that BEL classifies contract employees working off-site at another business as a separate establishment. This approach seems atypical of the standard definition which links employees who work out of a particular establishment back to that establishment rather than defining a new establishment for their temporary job

site. With this approach, a business could have as many establishments as they have employees, yet not have a physical space except perhaps at the central office. The locations at which their staff works as on-site contract labor might tend to be volatile over time.

Coordination is another issue that arises for administrative databases such as the BEL. The Bureau of Labor Statistics does not control the legislation under which these data are collected, nor do they have direct control over the State Employment Security Agencies who collect much of BEL data. The quality issues raised by this shared data collection and the extent to which state-to-state differences occur should be investigated.

Continuing business-based data collection can achieve gains in efficiency and data accuracy by using automated electronic methods of data collection. BLS is moving in this direction, and in doing so is creating or encouraging approaches that reduce respondent burden.

The Standard Statistical Establishment List

Walker (1997) describes the second contender for the title of the "Federal Government's Best and Most Complete Business Register." The U.S. Bureau of the Census uses tax return data to build the Standard Statistical Establishment List augmented with (1) "birth" data from the Social Security Administration [based upon applications for Employer Identification Numbers (EIN)] and (2) SIC codes for unidentified businesses from the BLS. SSEL is maintained through a register proving survey, the Company Organization Survey, and information derived from the Bureau's economic censuses and surveys. Use of these multiple data sources adds complexity to the operations and has an associated time delay attached to them. However, maintaining an accurate business register would not be possible without their use. Even with multiple data sources used, noticeable deteriorization in data quality occurs in the five years between the Company Organization Surveys. It was unclear how or if the Bureau uses sample-derived data to update the SSEL. Use of sample-derived data can bias future samples selected from the register, depending upon the circumstances.

An evaluation of SSEL data quality is now being planned, and I strongly encourage such action as many important economic surveys are derived from the SSEL. Walker also notes that, "Reconciliation of these business registers (the BEL and SSEL) at the level of individual statistical units would, no doubt, strengthen both registers and the Federal statistical system in general; unfortunately, confidentiality restrictions presently do not allow such effort." I concur with this observation that both registers would be enhanced by cross comparisons. The Federal statistical community needs to work to see the removal of such wasteful, artificial barriers to data quality and productivity.

Dun's Market Identifiers File

Marker and Edwards (1997) discuss Dun's Market Identifiers (DMI) File, which is used to construct the sampling frame for private surveys as well as for federal surveys with sponsors other than BLS and the Census Bureau.[1] The DMI file is constructed from lists derived from many sources, the principal ones being the Dun & Bradstreet business ratings and telephone listings. The DMI file is maintained for sale to users interested in direct marketing as well as business surveys. Over the last decade or so, considerable effort has been expended in improving the coverage of the DMI file and that effort has bourne fruit. That DMI file coverage is "near 98 or 99 percent" as Marker and Edwards assert was not quite proven to my satisfaction, however. Deadwood and duplicate records make their direct comparisons of frame counts problematic to interpret.

Out-of-business operations are not an uncommon occurrence in the DMI file. In part, such deadwood appears in the file because Dun & Bradstreet does not have access to good mechanisms for identifying business deaths. Rather, an absence of confirming evidence of life is observed by DMI operations. Once a firm goes out of business, the telephone listing may cease to exist, for instance. Quite rightly, in the absence of positive information as to death, DMI tends to retain the business record.

With known deadwood in the DMI file, comparisons of DMI frame counts to BEL or SSEL derived counts can be misleading. Such comparisons might be better made using weighted estimates derived once the sample has identified out-of-business operations. To the extent possible, the counts derived for comparison purposes should also be confined to types of businesses represented in both files.

---

[1]The Bureau of Economic Analysis gets access to SSEL through the 1990 Foreign Direct Investment and International Financial Data Improvements Act. Federal agencies can use the BEL after obtaining permission from the individual states.

Comparisons should be made at the enterprise and at the establishment level when possible and ideally should include breakdowns by business activity (manufacturing, transportation, service, etc.), ownership type (corporation, partnership, sole proprietorship), and size (number of employees).

The mechanisms used to build the DMI file also affect the quality of the individual DMI data. The number of employees variable is not reliable enough to use to exclude ineligible businesses. DMI-derived SIC codes show more correlation to survey-derived SIC codes but are often in disagreement, although the company usually reports a related industry type to that of the DMI. I recommend caution in using DMI data values to eliminate out-of-scope businesses. Incorrect entries can lead to sample undercoverage when database items are used to automatically exclude businesses from the sample. It may be preferable to include potential out-of-scope businesses in an initial screening interview which can also update location information and current operating status.

As did Searson and Farmer for BEL, Marker and Edwards compare coverage in terms of employment counts. This makes sense for BEL as BLS is interested in employment data. This comparison may also make sense for the National Employer Health Insurance Survey that Marker and Edwards mention, because here employee results are clearly of interest. However, larger businesses account for a substantial percentage of total employment, so coverage results for number of employees can be expected to be higher than for the number of businesses covered. Employee-based coverage rates can be deceiving for statistics unrelated to business size or surveys that collect such statistics. Examples might be a statistic such as the percent of businesses offering health insurance benefits or similarly a survey interested in the distributional attributes of businesses.

Some businesses may never be added to the DMI database. It is not uncommon for small businesses to be born and to die within a year or so. Marker and Edwards fail to factor in this consideration in projecting how long it takes before new businesses enter the DMI database.

My previous experience with the DMI file led me to resolve not to use the establishment-level data for sampling purposes. I was pleased that Marker and Edwards suggested that coverage at the establishment level has improved. I would still recommend that surveys sample at the enterprise level whenever possible, as DMI coverage of enterprises is more complete, almost by definition, than its coverage of the individual establishments each enterprise contains.

Concluding Remarks

It is sad that we have three different business registers being used for Federal business surveys. A friend of mine, Michael Colledge at the Australian Bureau of Statistics, calls this lack of data sharing between Federal statistical agencies--"that distinctly American problem." The Federal statistical community is making some progress in promoting data sharing, but the pressure should continue to remove these unnatural impediments to data quality and efficiency.

References

Marker, David A., and W. Sherm Edwards (1987). "Quality of the DMI File as a Business Sampling Frame." Invited paper, Joint Statistical Meetings, Anaheim, CA.

Searson, Michael A., and Tracey E. Farmer (1997). "Quality of the Bureau of Labor Statistics' Business Establishment List as a Sampling Frame." Invited paper, Joint Statistical Meetings, Anaheim, CA.

Walker, Ed (1997). "The Census Bureau's Business Register: Basic Features and Quality Issues." Invited paper, Joint Statistical Meetings, Anaheim, CA.