ESTABLISHMENT SURVEYS FROM THE END USER PERSPECTIVES[1]

Robert H. McGuckin, U.S. Bureau of the Census
Robert H. McGuckin, Bureau of the Census, CES, Room 1587-3, Washington, D.C. 20233

KEY WORDS: Establishment, Access, Microdata, Statistical Purpose, Regulatory Purpose

I. Introduction

Establishments are physical locations that carry out the activities (production, distribution, central coordination, research and development, etc.) of firms. Firms own (or contract for) one or more of their locations in carrying out their operations. In this paper, I discuss the importance of longitudinal establishment microdata linked across surveys and censuses and tied together at each point in time by ownership identifiers. Establishment data are also the best source of data for forming many aggregates. For example, local area aggregates cannot be formed from geographically diversified firm-level data. This and similar uses of establishment data are not emphasized here. In fact, most of the discussion considers the limitations of aggregate data.

I also consider the implications of cross-survey linkages of establishment data for sharing data among statistical and regulatory agencies and the need for developing flexibility in arrangements for user access to the confidential microdata. Since much of my discussion relies on studies conducted at the Center for Economic Studies (CES) at the U.S. Census Bureau, whose work depends on one form of flexible access arrangements, I begin with the access issue.

II. User Access to Microdata is Possible and Desirable

In a typical year, the U.S. Census Bureau, like statistical agencies around the world, carries out numerous surveys or censuses of firms and establishments. These data are routinely processed, tabulated, and published as cross-section aggregates -- the minimum level of aggregation being determined by considerations of confidentiality. The effort is directed toward producing as much tabular output as possible with emphasis on the production of aggregate point-in-time cross-section estimates for use in the National Income and Product Accounts (NIPA). These estimates include input-output tables, price indexes, and productivity measures for various sectors. Once these tasks are completed, little in the way of resources has been devoted to making further use of the basic data. From the operations point of view, the microdata are expendable once the tabulations are made. Moreover, the Office of Management and Budget (OMB) and other oversight agencies have generally not put much priority on analysis of microdata. Reasons for this are discussed in Triplett (1992). Therefore, documentation and storage have not been undertaken with care. These incentives to ignore the historical data have been reinforced by budget stringencies over the past 10 to 15 years.

The CES program begins where the regular data production programs leave off. CES does not conduct surveys of its own. Instead, survey microdata collected as part of regular programs are linked to form longitudinal panels at CES and these panels are also broadened by linking them with other microdata -- both survey and administrative -- from both within and outside the U.S Census Bureau.

Business microdata panels were virtually non-existent just 10 years ago. Today, however, CES creates and maintains panel datasets at the establishment, firm, and enterprise level; further, there are a large number of similar efforts underway and in the planning stages around the world. See the Proceedings of the First Annual Eurostat Conference on Longitudinal Panels (1994).

Closely related to the CES panel data creation and maintenance program is a plan that provides researchers from government, academic, and private research organizations with direct access to these microdata panels. Each year, between ten and twenty outside researchers become Special Sworn Employees (SSEs) of the U.S. Census Bureau and work as research associates at CES. With the opening (1/4/94) of a CES regional facility -- the Boston Regional Data Center (BRDC) -- the number of researchers per year should increase by 30 to 40 percent. This access program has led to a wide array of influential academic and policy studies by leading researchers and CES staff. The wealth of information being developed in this program is providing new insights into everything from job creation and destruction to the effects of enterprise ownership changes on productivity and jobs. See CES Annual Reports (1990-94), McGuckin and Reznek (1993), and McGuckin (1992, 1994).

III. Why Establishment Data Are Important

Economic analysis is generally based on a model of individual behavior that specifies the objectives and constraints facing the economic agent. Agents are assumed to maximize an objective function subject to those constraints. The maximization problem can be specified either as a static or dynamic optimization. Solutions to such problems provide relationships for the endogenous variables -- the variables the agent has choice over -- as functions of exogenous variables of the model. That is, the model specifies the functions that determine certain decisions the agent makes (e.g., using a particular type of fuel in a manufacturing process or producing cars with particular characteristics.) Using estimates of the parameters of these models, analysts examine the effect of some change in the agent's environment (the factors that affect decisions) on her behavior.

For many problems, the establishment is a sensible unit of analysis. For example, from the standpoint of the production decision, the choice of labor, energy, materials, and capital for use in output creation is often made at the plant level. While the firm is the ultimate decision maker, and thus the preferred unit of analysis for many problems, our data shows that establishments often have very different behavioral patterns, even when owned by the same firm. Thus, establishment data are also necessary in order to understand the behavior of the firm. (The behavior of one establishment is not completely differentiated from another simply by the identity of its owner.) Establishment data are also necessary to estimate the marginal impact of some event -- for example, a purchase or divestiture of assets -- on the firm. Moreover, for multi-establishment firms -- the most important in terms of, for example, employment or output -- the establishment-level data is required for analysis of the impact of policies and actions on particular locations or regions.

Identifying the Production Relationship

Focusing on the production relationship, one can see that the establishments are the primary purchasers of the factors of production: labor, materials, capital services, and energy. Even though primary resource allocation decisions are often made at the firm level, establishment data are useful in the analysis of technical change -- both product and process -- since technical progress is characterized by changes in the production relationship that manifest themselves at the plant level. Similarly, the establishment must be the unit of analysis for many environmental issues. Environmental problems involve the production of two outputs, products (good output) and emissions (bad output), at the same establishment. Understanding the relationship between the "good" and the "bad" output is essential in developing environmental policies.

Measuring the Effects of Changes in Ownership

Another area of study in which establishment microdata are essential is in the evaluation of the effects of ownership change (mergers, divestitures, leveraged buyouts, etc.). Firms, particularly large ones, have multi-establishment structures, and many firms are diversified across a wide spectrum of industries and products. To assess the effects of mergers, the analyst must separate out the components of the firm both at a point in time and across time. This allows for evaluation of the performance of the firm and its components pre- and post-merger. A number of recent studies have found significant impacts of ownership change on establishment productivity. See Lichtenberg (1992), Long and Ravenscraft (1992a,b), McGuckin and Nguyen (1993), and McGuckin, Nguyen, and Andrews (1991).

Aside from policy issues, studies of ownership change are essential to improve our understanding of the evolution of the establishment and firm sampling frames used to generate typical aggregate cross-section economic statistics.

Examining the Structure and Boundaries of the Firm

The use of establishment data in understanding the nature of the firm is not restricted to analysis of the role of ownership change in performance. Since the establishment is often the site for particular sets of products and processes, it provides the natural unit for understanding the nature and interrelationships of the activities of the firm. As a geographically fixed production unit, the establishment makes it possible to identify the role of "the firm" -- as distinct from such factors as geographical and product markets -- in the establishment's operating characteristics and behavior. In turn, this provides insight into the functions and boundaries of the firm.

For example, Streitwieser (1991) shows that the pattern of secondary products produced by establishments owned by the same firm is much more closely related than those produced by other establishments with the same primary products. As another example, Gollop and Monahan (1991) find that while the structure of production has become more specialized over time, the structure of the activities of the firm have become more diversified. While the

focus of these inquiries is the nature of the firm, the data required are from the establishment.

Understanding The Labor Market

The importance of explicitly dealing with establishments as economic agents is also illustrated in the literature on labor markets. Until very recently, labor market analyses were carried out almost exclusively by using microdata on household or individual workers -- the supply side of the labor market. This type of research still represents the vast majority of such studies. Yet economic theory suggests that the demand side of the market is no less important than the supply side. Several recent studies at CES illustrate the importance of the demand side of the market.

An important area of labor market research involves attempts to explain earnings differentials among individuals by various characteristics such as education, sex, race, age, family status, and occupation. These studies offer much insight into the factors that explain differences in earnings and have been important in formulating social policies (e.g., support for education, a factor that is positively related to earnings). Despite a large literature on this subject, including analyses of available public use microdatasets on individuals, and despite enormous interest by economists, social planners, sociologists, and policymakers, among others, the earnings models explain less than 50 percent of the variance in earnings in most studies.

One explanation for this is that it has been difficult to include the demand side of the market in earnings equations. Studies at CES have documented important differences in wages associated with individual establishment characteristics. For example, in Dunne and Schmitz (1992), plants adopting advanced technologies are found to pay higher wages than those that do not. This suggests that these plants may hire highly skilled workers -- workers with skills that may not be captured completely by demographic variables such as education and experience. Further, factors such as on-the-job training and learning by doing contribute to earnings and are, at least partially, reflected in the characteristics of the plant where the worker works.

The possibility that plant-level characteristics would improve earnings equations was one reason that CES developed an experimental database linking workers to individual plants. While the current sample is not representative of the entire population of establishments or workers, preliminary analysis has yielded several key insights. See Troske (1993, 1994) and Doms, Dunne, and Troske (1994).

Aggregation Issues and the Microfoundations of Macroeconomics

Earlier I noted that most published data reported by the U.S. statistical system are aggregations. These aggregations reduce the myriad of individual detail to manageable proportions and provide confidentiality protection. Unfortunately, information is lost or distorted in this aggregation process. For some problems, this loss of detail may not matter: the phenomena under study may be sufficiently understood without reference to the underlying microdata. Without analysis of the microdata, however, it is virtually impossible to evaluate the extent of any aggregation error.

As an example of the importance of taking account of this establishment heterogeneity, consider the typical response to the fact that the manufacturing sector has increased productivity and has shed workers over the past 10-15 years. The conventional wisdom on this is that rising productivity in the manufacturing sector is due to firm downsizing. But the evidence suggests that this is a misleading picture. In 1987, "upsizing winners" (manufacturing plants that increased both employment and labor productivity between 1977 and 1987) accounted for 39 percent of total manufacturing employment and 43 percent of output, while "downsizing winners" (plants that decreased employment and increased productivity) accounted for 28 percent of employment and 35 percent of output. See Baily, Bartelsman, and Haltiwanger (1994).

This project is part of a major new CES research program to investigate the microfoundations of macroeconomics. This research is at the center of an important recent development in macroeconomics: the idea that understanding aggregate fluctuations requires analysis of time-series fluctuations in the cross-sectional distribution of activity across establishments. For example, the conventional view of recessions -- that jobs disappear temporarily while the creation of new jobs declines, and that most workers are recalled when aggregate demand recovers -- appears incorrect (at least in the manufacturing sector). In fact, job creation continues almost at the same pace during recessions, while job destructions increase. Furthermore, most jobs created are created permanently, and most jobs lost are lost permanently. See Davis and Haltiwanger (1990, 1992) and Davis, Haltiwanger, and Schuh (1994). These facts suggest that the standard empirical approach to business cycle analysis (based on representative agent macroeconomic models), which assumes that firm behavior is symmetrical over the cycle, is incorrect. This research also shows that

variations in new job creations and destructions, which can be calculated only from longitudinal microdata on establishment employment, are primarily associated with movements among plants within the same industry. That is, both lost jobs and new jobs are observed simultaneously in the same industry as transfers from one plant to another. This means that, for example, the effects of regulatory changes that may force firms to substitute away from labor and toward capital in production will depend on the detailed characteristics of the distribution of plants within an industry and cannot be captured by a representative or average industry response.

Aside from its policy relevance, this new line of macroeconomic research has fundamental implications for statistical data programs. One implication is the clear need to develop longitudinal panels and make job creation and destruction statistics a part of the regular statistical programs for the entire economy. A second is the expansion of user access programs so that the to allow analytic users -- researchers and academics -- to explore composition effects and carry out effective benchmarking studies. Finally, the job creation and destruction studies carried out at CES illustrate an important point: access programs can have important benefits to the statistical agency in terms of the quality and relevance of their data products.

New product generation, as is illustrated by the job creation and destruction work at CES, is of particular importance. As I have noted, this research showed that aggregate measures of the mean of the distribution of economic activity within sectors are inadequate for macroeconomic research. One possibility for improving economic data is the construction of new measures of economic activity based on higher level moments (e.g., variance, skewness, and kurtosis) of the distribution of, for example, aggregate output. This latter possibility is suggested in work by Caballero (1992), Caballero and Engel (1992), and Haltiwanger (1993) and is the subject of ongoing research.

These examples illustrate that we need to do more than just evaluate aggregation bias in traditional models. The problem with exclusive use of aggregate statistics is not simply one of inferior estimates of traditional economic relationships such as the earnings equation, elasticities of production functions, or inventory adjustment coefficients. The fundamental problem is that with aggregate data alone, it is impossible to examine the differential effects of policies on the entities (establishments or firms) classified within the aggregate. Longitudinal microdata are necessary to sort out the fundamental roles of entry, exit, and changes (growth and decline) in establishments. In other words, examination of individual changes is necessary if particular components of an aggregate movement are significant. For this, a program of user access to establishment microdata is necessary.

IV. Cross-Survey Data Linkages and Data Sharing

While we have not emphasized it to this juncture, many of the research projects at CES involve linkages between different surveys. For example, earlier we noted several CES studies using a data set that matched worker information from the 1990 Decennial Census with establishment production data from the Longitudinal Research Database (LRD) at CES.

A substantial number of economic and policy problems require that data from a variety of sources be linked. Moreover, the linkages usually require access to micro establishment data, as shown in the following examples.

Not all plants classified in particular SIC industries are unionized. Therefore, to assess the performance of union and nonunion establishments, the individual establishment data are required.

Another example of important work relying on linked microdata is a new training survey being conducted by the U.S. Census Bureau under contract with the Department of Education. An explicit goal of the study using this survey is to measure how training affects establishment performance. The important new wrinkle in this survey is that the sample design explicitly ensures that we will be able to match the survey data on training with longitudinally-based measures of establishment performance from the LRD, such as productivity. This design reduces the need for new survey information on establishment performance to provide a benchmark for evaluating training expenditures.

There is also provision for researchers from the sponsoring agency to access the microdata directly at the new BRDC. This part of the design will provide for more analytical work with the data and extend the information available to the users beyond that which could be obtained from tabulations that preserve confidentiality.

All this is to suggest that the time for more extensive sharing of information among statistical agencies and with users is here. Sharing among statistical agencies would prevent duplication of effort and most likely reduce respondent burden. It would allow for coordination among agencies in the design of surveys and, like the training survey discussed above, provide more "bang" for the survey dollar. In other words, the benefits from traditional data programs would increase substantially at very small cost.

While calls for such sharing are not new, sharing among statistical agencies alone will not do. The sharing must also extend to analytical users -- for projects with scientific integrity that are empirical in nature. As discussed in some detail elsewhere (Triplett 1992, McGuckin 1992, 1993, and the National Science Foundation grant for the BRDC) access to the confidential microdata by analytical researchers is crucial for both users and the statistical agencies.

The record of CES stands clearly for the proposition that the heterogeneity among establishments cannot be ignored if one wants to make correct policy choices. It also illustrates the value of analytical research in developing data products and sound conceptual designs for surveys.

V. Concluding Comments

It is possible to create matched datasets at an any level of aggregation if there is a common classification scheme in each dataset (e.g., a SIC code) but, for the reasons discussed above, this is not the best procedure. While it is impossible to precisely quantify the size of the errors or limits to conclusions drawn from such matched aggregates, CES research has shown that they are probably large. The large degree of heterogeneity in establishment characteristics within common sectorial classifications such as industry, region, and size class almost guarantees it. See Davis, Haltiwanger, and Schuh (1994). It is not just heterogeneity, of course, which is at issue. What is striking is that most variation in the observable characteristics -- job creations and destructions, investment, for examples -- is associated with within sector changes, irrespective of the aggregation level.

In closing, I want to emphasize several aspects of sharing that I raised in a panel discussion at last year's American Statistical Association meetings. As discussed above, with proper confidentiality restrictions, information should be broadly available to researchers within the statistical system. I include in this administrative data collected by the Internal Revenue Service and other regulatory agencies, data collected by the statistical agencies, and public databases that are often a part of research projects undertaken by analytical researchers. In fact, administrative data is already used in the CES program and is crucial to almost all statistical programs at the U.S. Census Bureau.

It is also important to recognize that individual confidentiality must be protected -- not only because without such protection it is likely that the public will not cooperate with requests for information, but also because it is morally and legally unethical to use such data for "regulatory" purposes.

But what do we mean by regulatory purposes? In my view, all output of statistical systems supports regulatory purposes. Even Gross National Product statistics, which most people would not view as supporting regulatory purposes are an important influence on policy and regulation. Where do we draw a line? In my opinion, any output of a statistical agency -- analytic study or tabulation -- that meets confidentiality standards should be deemed created for "statistical" not regulatory purposes. (I am aware that deciding on exact confidentiality standards is not a simple matter. However, I think there is a long history of successfully dealing with the problem, even in regards to user access. See McGuckin and Nguyen (1991) and McGuckin (1992). In essence, regulatory purposes would be isomorphic with confidentiality standards -- anything that breached individual confidentiality would be deemed as having been created for regulatory purposes. This would eliminate the confusion surrounding the term "regulatory purposes" and minimize mindless legalistic arguments.

This proposal would also improve policymaking. Statistical agencies and analytic users would have rich datasets to develop at the lowest possible cost. These datasets would improve both traditional aggregate data products and analytic studies that provide policy guidance in a general way. Published tabulations or analytic study results would not identify individual persons or businesses directly, only as part of a broad class such as African-American business owners, exporters, plastic manufactures, etc.

With data sharing and user access, (under carefully controlled conditions) policy-relevant information would be available to the regulatory authorities and law enforcement agencies, as well as private organizations and the Congress. Such information is likely to be better than available now for two reasons. First, studies will rest on a broader and better base of information because of the cross-survey and cross-agency data from linked databases developed at the statistical agencies. Second, the analytic researchers will be able to develop more precise estimates and examine composition effects by taking advantage of the microdata.

REFERENCES

Baily, M.N., E. Bartelsman, and J. Haltiwanger (1994), "Downsizing and Productivity Growth: Myth or Reality?," mimeo, Table 2, p. 13.

Caballero, R.J. (1992), "A Fallacy of Composition," American Economic Review.

Caballero, R.J. and E.M.R.A. Engel (1992), "Beyond the Partial Adjustment Model," American Economic Review, Papers and Proceedings.

Davis, S.J. and J. Haltiwanger (1990), "Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications," National Bureau of Economic Research Macroeconomics Annual, MIT Press, pp. 123-168.

_____ (1992), "Gross Job Creation, Gross Job Destruction and Employment Reallocation," Quarterly Journal of Economics (Vol. 107), pp. 819-864.

Davis, S.J., J. Haltiwanger, and S. Schuh (1994), Gross Job Flows in U.S. Manufacturing, U.S. Department of Commerce, Bureau of the Census, Center for Economic Studies.

Dunne, T. and M.J. Roberts (1993), "The Long-Run Demand for Labor: Estimates from Census Establishment Data," Center for Economic Studies Discussion Paper 93-13.

Doms, M.E., T. Dunne, and K. Troske (1994), "Workers, Wages, and Technology," CES working paper.

Dunne, T. and J. Schmitz (1992), "Wages, Employer Size-Wage Premia and Employment Structure: Their Relationship to Advanced-Technology Usage at U.S. Manufacturing Establishments," Center for Economic Studies Discussion Paper 92-15.

Gollop, F. and J.L. Monahan (1991), "A Generalized Index of Diversification: Trends in U.S. Manufacturing," Review of Economics and Statistics (Volume 73, 2), pp. 318-330.

Haltiwanger, J. (1993), "Proposal for Census Fellows Program".

Lichtenberg, F.R. (1992), Corporate Takeovers and Productivity, MIT Press.

Long, W.F. and D.J. Ravenscraft (1992a), "LBOs, Debt, and R&D Intensity," in Decade of Debt: Lessons from LBOs in the 1980s, M. Blair, ed., Brookings Press, Washington, D.C.

_____ (1992b), "The Financial Performance of Whole Company LBOs," Center for Economic Studies Discussion Paper 93-16.

McGuckin, R.H. (1992), "Analytic Use of Economic Microdata: A Model for Researcher Access with Confidentiality Protection," Proceedings, International Seminar on Statistical Confidentiality.

_____ (1993), "Multiple Classification Systems for Economic Data: Can A Thousand Flowers Bloom? And Should They?," Journal of Economic and Social Measurement, (Volume 19), pp. 179-198.

_____ (1994a), "The Creation and Use of Microdata Panels: Insights From the Center for Economic Studies' Experience," forthcoming, Proceedings of the 1st Eurostat International Workshop on Techniques of Enterprise Panels.

_____ (1994b), "Establishment Microdata for Economic Research and Policy Analysis: Looking Beyond the Aggregates," forthcoming, Journal of Business and Economic Statistics.

McGuckin, R.H. and S.V. Nguyen (1991), "Public Use Microdata: Disclosure and Usefulness," Journal of Economic and Social Measurement, Volume 16, Number 1, pp. 19-90.

_____ (1993), "On Productivity and Plant Ownership Change: New Evidence from the LRD," forthcoming, RAND Journal of Economics, Center for Economic Studies Discussion Paper 93-15.

McGuckin, R.H., S.V. Nguyen, and S.H. Andrews (1991), "The Relationships Among Acquiring and Acquired Firms' Product Lines," Journal of Law and Economics, (Volume 2), pp. 477-502.

McGuckin, R.H. and A.P. Reznek (1993), "Research With Economic Microdata: The Census Bureau's Center for Economic Studies," Business Economics, (Volume 28,3), pp. 52-58.

Olley, G.S. and A. Pakes (1992), "The Dynamics of Productivity in the Telecommunications Equipment Industry," unpublished working paper.

Streitwieser, M.L. (1991), "The Extent and Nature of Establishment Level Diversification in Sixteen U.S. Manufacturing Industries," Journal of Law and Economics, (Volume 2), pp. 496-551.

Triplett, J. (1992), "The Federal Statistical System's Response to Emerging Data Needs," Journal of Economics and Social Measurement, Volume 17.

Troske, K. (1993), "The Worker Establishment Characteristics Database," mimeo.

Troske, K. (1994), "Evidence on the Employer Size-Wage Premia From Worker-Establishment Matched Data," mimeo.

U.S. Bureau of the Census, Center for Economic Studies (1990-1994), "Center for Economic Studies Annual Report".

U.S. Bureau of the Census, Center for Economic Studies (1993), "Proposal to the National Science Foundation for the Boston Regional Data Center".

END NOTES

1. This paper is for "Establishment Surveys: Where We Are Now -- Where We Should Be Going," invited panel discussion, 1994 Joint Statistical Meetings in Toronto, Canada, August 14-18, 1994. The opinions and conclusions expressed in this paper are those of the author and do not necessarily reflect those of the U.S. Bureau of the Census.