7-1-2008

Comparative Analysis of Enterprise Data (CAED): A Research and Data Agenda

John S. Earle
W.E. Upjohn Institute for Employment Research

John S. Earle

Comparative Analysis of Enterprise Data (CAED)
A Research and Data Agenda

Ten years ago, an economist leafing through the major professional journals would have been hard-pressed to find many articles using firm-level data. Particularly unusual were studies using comprehensive panel data on all enterprises in a single economy, and still rarer, practically unknown, were analyses of such data for multiple countries. One of the most important developments in economic research over the last decade is the growing analysis of such databases. The new data provide the opportunity for revisiting many of the classic empirical questions in economics, this time with data at the appropriate level of aggregation—the business that is the decision-making unit. The data also permit and stimulate the analysis of many new questions that economists could not even dream of addressing with previous data resources. Together, the data and accompanying research agendas are transforming much of economics and public policy analysis.

The Upjohn Institute has contributed to these developments both through in-house research and by partnering with other research and policy groups to organize a recent international conference and a new research network including economists and statistical agency officials from around the world. This article provides a brief, selective overview of the new types of data and research, and then discusses the Institute's organizational efforts, in particular the Conference on Comparative Analysis of Enterprise Data (CAED) and the research network.

New Types of Enterprise Data

While economists have studied firm-level data sets for a long time, the quantity, quality, and availability of the data have all vastly expanded in recent years. Earlier data sets on businesses tended to be small sample surveys focused on specialized topics and containing only cross-sectional information. Individual researchers frequently assembled these data on their own, or they organized the collection for the purpose of a particular research project. Limited funding generally resulted in small-scale data sets, little standardization of variables, and little sharing of data among researchers. Moreover, despite the usefulness, indeed the necessity, of such data for answering a variety of questions, the tendency of the economics profession to award little credit for data collection meant that most economists felt only weak incentives to expend effort in this area. It was much easier to work with standard, existing databases on households or industry, regional, and economy-wide aggregates.

The new data sets on businesses tend to rely on governmental sources, and as a consequence they are more systematic and much larger in scale. Both the national statistical offices and the agencies administering government programs have regularly collected data on firms and establishments in order to monitor the macroeconomy, collect taxes, and evaluate policies. But researchers were unable to obtain access to the business-level information. A number of recent developments—growing openness of governmental agencies, increasing pressure from the research community, improving technologies to process data and protect confidentiality, and mounting emphasis on empirical research, particularly at the micro level—have led to accelerating access and analysis of the microdata.

The new data sets have several important advantages. Numbers of observations are much larger, permitting stronger conclusions from a given

analysis. In some cases, the data are "universal," covering the entire population of businesses in a country or a particular sector (all of manufacturing industry, for example). In other cases, these databases make use of universal sampling frames, which solves one of the biggest problems confronting researchers carrying out their own firm surveys: constructing a representative, random sample.

The new data are also usually longitudinal, containing multiple observations over time for a given business. This panel dimension of the

---

The quantity, quality, and availability of firm-level data have all vastly expanded in recent years.

---

data facilitates statistical techniques to control for unobserved differences across firms, and it permits researchers to study dynamics—changes in firm behavior and responses to shifts in the firm's operating environment. When combined with universal coverage, a dynamic analysis can also address questions involving entry and exit of businesses from the market. The dynamics of firm turnover are not only of great potential consequence for economic growth and worker welfare, but they also may be important statistical factors to control for, as analyses based on only continuing firms are likely to be biased if exit or entry is nonrandom.

A significant drawback of the new data sets is the limitation to a relatively small set of variables used for constructing aggregates and evaluating programs. Moreover, while the data have become much more readily available for researchers than they were in the past, large obstacles to access persist in many countries. Cross-country comparisons are facilitated by some degree of standardization in the collection of data for national income accounting, but idiosyncrasies in definitions of variables and in the rules for inclusion of observations (in the sample or universe) remain. Thus, individual surveys focused on particular topics will

continue to play an important role, as will the collaboration of researchers with knowledge of local idiosyncrasies in data, policies, and institutions with the data providers to link, harmonize, improve, and make available many types of data.

One important subclass of enterprise data sets deserves particular mention: linked employer-employee data (LEED). Such data contain information on the composition of employment within firms, including the characteristics and, usually, the wages of workers. The information is useful for controlling for differences across firms in the workforce and for studying many questions involving the internal organization and compensation structures of firms. When the data contain longitudinal information on both employers and employees, it is also possible to control for unobserved firm attributes in analyzing worker outcomes and to study job mobility of workers across firms. In essence, the data permit analysis of both the demand and the supply sides of the labor market.

New CAED Research

The new data permit many of the fundamental questions in empirical economics to be studied at the level of the enterprise, the decision maker for many questions underlying economic growth and welfare. In traditional economics, for example, the entire economy (or an entire industry) is modeled as if it were a single firm with a single production function it uses to transform inputs into output. With the aggregate data corresponding to such a model, estimation of basic parameters is at best difficult, because sample sizes are too small to permit reliable inference. More importantly, the assumption of a common technology across diverse industries is untenable. To take one example, in projecting the impact of a rise in oil prices on employment levels, the researcher needs to estimate cross-elasticities of input demands, which depend on technological ease of substitution and market factors that vary across industries. Using industries as observations to estimate these relationships for the economy as a whole fails because the industry and

economy-wide aggregate relationships are in general different, so no inferences are possible. These problems can only be avoided by moving to the firm level, using the data corresponding to the decision maker, and estimating separately by groups operating in relatively homogeneous markets and with homogeneous technology.

Many important economic questions cannot even be empirically posed in the absence of appropriate enterprise data. One of the earliest lessons from firm-level research is that firms display enormous heterogeneity in their performance and behavior, even within narrowly defined categories and industries. The diversity of outcomes contradicts standard theoretical models of competitive industries and frictionless environments as well as empirical analyses based on aggregate (sectoral or regional) data. The factors leading some firms to be more productive than others are fundamental determinants of economic growth, and they are fundamental puzzles for economists, but they can only be satisfactorily investigated with firm-level panel data. Among the factors that the new literature

---

The new data permit many of the fundamental questions in empirical economics to be studied at the level of the enterprise, the decision maker for many questions underlying economic growth and welfare.

---

is addressing are technology and R&D, ownership and corporate governance, and government policies and institutions.

Another set of questions that requires enterprise data, ideally with universal coverage, concerns industry dynamics. Stretching back to Schumpeter, there has been much casual discussion of the potentially important role played by the creative destruction process in capitalist economies. But the data required to investigate the nature of exit and entry have only recently become available. The important research issues concern the pace and the determinants of the

firm turnover process, as well as its consequences for growth: the shares of entrants and exiters, their relative productivity levels, and the magnitudes of costs of entry, exit, and remaining in the market. Closely related policy questions involving these costs include regulatory barriers to entry, financial constraints on growth, competition from entrants and international trade, provision of complementary institutions, and softness of budget constraints.

A final set of issues involves the consequences of the firm-level restructuring and reallocation processes for workers. Do employees gain when their employer's productivity improves? Or does the improvement more often come at their expense? These questions can be addressed in the context of any of the factors or policies affecting firm performance and industry dynamics. Using firm-level data, the outcomes for levels of employment and average wages of the firm may be estimated. Using LEED, it is possible to estimate heterogeneous outcomes for different types of workers and, in some cases, to trace the mobility patterns and long-term consequences for displaced workers.

Ultimately, the analysis of firm-level data promises new insights into the

The Upjohn Institute has contributed both through in-house research and by partnering with other research and policy groups to organize a recent international conference and a new research network including economists and statistical agency officials from around the world.

causes of differences in the "wealth of nations." Most firm-level research has been carried out for single countries, but many of the important decisions underlying international differences in productivity and incomes reside within firms. International comparative research on enterprise data can exploit variation in policies and institutions and their effects on enterprise behavior, thus revealing the

reasons for cross-country differences in firm performance and industry dynamics.

The CAED Conference and Network

In order to bring together scholars and civil servants working with the many strands of research and types of enterprise data, the Upjohn Institute recently organized an International Research Conference on Comparative Analysis of Enterprise Data in partnership with the Central European University in Budapest, Hungary. The conference was the 8th in a series of CAED conferences that emerged from workshops at the U.S. Department of Commerce in the mid-1990s and have since been held in several countries. The major sponsor of the Conference was the COST (Cooperation in Science and Technology) program of the European Science Foundation, and other cosponsors included the Hungarian National Bank, the Institute for Employment Research (IAB, Nuremberg), the Hungarian Competition Authority, the U.S. Census Bureau, the National Opinion Research Center, and the Organisation for Economic Co-operation and Development.

The conference inaugurated a new international network organized by the Upjohn Institute and the Central European University with the support of a four-year grant from COST. The purpose of the network is to bring together leading researchers from around the world to work with national statistical agencies and to collaborate on new cross-country comparative research investigating the roles of industry dynamics and firm performance in economic growth as well as their consequences for employees. The network will organize workshops and conferences following the lines of the research initiatives discussed above: industry dynamics, firm performance, and worker outcomes. In addition, a special working group will focus on issues of data access and quality, which are relevant for all researchers in this area.

The Upjohn Institute will continue to play an active role in CAED both by contributing research and by helping the network to expand to a wider range of countries and economists. The Budapest

conference already brought together researchers and data providers from 26 nations, but the CAED research agenda would clearly benefit from comparative analysis based on a larger and more diverse set of policies and institutional experiences. The network will also

Using LEED, it is possible to estimate heterogeneous outcomes for different types of workers and, in some cases, to trace the mobility patterns and long-term consequences for displaced workers.

encourage work on firm-level data within many different fields of interest and by a variety of types of economists. Some fields, including labor economics, industrial organization, and international trade, have been quick to incorporate firm-level data, but many more stand to benefit, as do researchers in other social sciences. Finally, the network will help foster the development of a new generation of researchers. Although competition for paper presentations at the Budapest conference was stiff (110 papers were accepted out of more than 260 responses to an open call), an unusual number of papers were coauthored by early stage researchers, including current graduate students. It seems safe to predict that CAED growth over the next 10 years will be even more rapid than in the past decade.

John S. Earle is a senior economist at the Upjohn Institute.