

This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: The Analysis of Firms and Employees: Quantitative and Qualitative Approaches

Volume Author/Editor: Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 978-0-226-04287-9; 0-226-04287-1

Volume URL: <http://www.nber.org/books/bend08-1>

Conference Date: September 29-30, 2006

Publication Date: October 2008

Chapter Title: Introduction to "The Analysis of Firms and Employees: Quantitative and Qualitative Approaches"

Chapter Author: Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, Till von Wachter

Chapter URL: <http://www.nber.org/chapters/c9110>

Chapter pages in book: (1 - 16)

Introduction

Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter

Motivation for Book

The analysis of the interactions of firms and employees has followed two distinct paths. One path has focused on large-scale, often nationally representative, data sets on firms and employees, typically housed at federal statistical agencies. In some cases, this path has intensively used administrative data, alone or integrated with survey data and, in other cases, the use of surveys designed to collect information about both firms and workers. The other path has been the development of specialized surveys and gathering of personnel records of a small number of firms (or even one firm) or intensive observation (essentially collection of qualitative data) from case studies based on site visits to firms by researchers, data typically housed at universities or think-tanks.

Each of these two study approaches has uncovered interesting and useful pieces of information. Researchers working with large-scale, national, firm-level matched employer-employee data sets have begun to address a variety of organizational topics, such as determinants of wage inequality, the use of alternative wage policies (such as the use of incentive pay) and their impact on worker selection, gender differences in promotion, and

Stefan Bender is a senior researcher at the Institute for Employment Research. Julia Lane is senior vice president of Economics, Labor, and Population Studies at the National Opinion Research Center at the University of Chicago, and a senior research fellow at the U.S. Bureau of the Census. Kathryn Shaw is the Ernest C. Arbuckle Professor of Economics in the Graduate School of Business, Stanford University, and a research associate of the National Bureau of Economic Research. Fredrik Andersson is a senior research associate of the Cornell Institute for Social and Economic Research, and a research fellow with the Longitudinal Employer-Household Dynamics Program (LEHD), U.S. Bureau of the Census. Till von Wachter is an assistant professor of economics at Columbia University, and a faculty research fellow of the National Bureau of Economic Research.

differences in alternative career paths within and across firms. They have also been able to examine the impact of job and worker reallocation on worker and firm outcomes. Still, one weakness of existing empirical analysis of firm outcomes from large, national data sets is that the underlying determinants of firm performance are generally unknown.

In contrast, researchers who have been engaged in intensive study of a small number of firms through either case studies or specialized surveys have been able to describe rich contextual variation in organizational decision making. Many of these studies have been conducted within the context of projects affiliated with the Sloan Foundation's Industry Centers and the NBER/Sloan joint projects sponsoring site visits. These studies have highlighted, among other things, the importance of human resource practices such as the use of teams for the successful adoption for the use of new technologies (e.g., information technology) and have as a common theme the link between business success how businesses organize their workers. While such work has demonstrated the tremendous value added of specialized surveys and the insights to be derived from the intensive qualitative data collection associated with site visits, questions are often raised about the ability to generalize results from small, potentially nonrepresentative samples.

The combination of these two empirical approaches means that it is possible to envision the development of a new field of economics, one that is at the nexus of labor economics, industry studies, and industrial organization. The focus of this field is inherently the organizational structure of businesses with a focus on how workers are organized. Matched employer-employee data that include the information from large-scale data sets as well as from specialized surveys and site visits provide a momentous opportunity for a research agenda that is focused on the study and understanding of the interaction of firms and workers.

This book results from a conference that was planned to foster just such a research agenda.¹ An overarching goal of the conference was to bring together both senior and junior researchers from the two study fields—traditional labor economists and industry studies researchers, particularly those who have conducted case studies—to illustrate the different insights to be gained from the two approaches and provide the stimulus for a next

1. The Conference on Firm and Employees (CAFE) was held September 29 to 30, 2006, in Nuremberg, Germany, sponsored by the Institute for Employment Research (IAB), the Data Access Center (FDZ-BA/IAB), The Deutsche Forschungsgemeinschaft (German Research Foundation), their Research Network "Flexibility in Heterogeneous Labour Markets," the National Bureau of Economic Research, the Alfred P. Sloan Foundation, and the National Science Foundation. Invited keynote speakers included Dan Hamermesh (University of Texas, Austin), Ed Lazear (Stanford University), Richard Freeman (Harvard University), and Mari Sako (Oxford University), who all have done path-breaking work related to the topics of the conference. Over 160 papers were submitted, of which about 40 were accepted after an extensive refereeing process. Over 100 researchers from around the world participated.

generation of research. The potential for such data to answer key empirical questions in economics has been well described elsewhere.² The list includes the effect of firm policies on the job ladders and career paths of workers; the effects of workforce composition on business growth and survival; as well as the micro-level analysis of the demand for labor, including the impact of technological and structural change.

The book features eleven papers selected from that conference. They were selected by referees on the basis of their quality as well as for the new insights that they provided about the interactions between firms and their employees. The next sections of this introduction provide an overview of the major findings that have been made possible as a result of these new data sets. These span several different facets of the relationship between firms and workers, beginning with shedding more light on the relationship between human resource practices and productivity, then examining how firm differences in the organization of production are related to differences in human resource practices, how changing ownership affects the organization of production, and, finally, how the changing trade patterns, particularly globalization, affect firm competitiveness and then works through to their employees. The final section provides an overview that highlights the innovative nature of the data sets themselves.

The Major Findings

Human Resource Practices and Firm Productivity

Personnel economics has long been concerned with the fundamental question whether human resource practices such as incentive contracts or monitoring affect workers' productivity and worker turnover and how they relate to firms' efforts to innovate and gain competitiveness in an evolving marketplace. A recent surge in new data sets and data collection efforts has led to an increasing amount of ambitious empirical work describing and testing some of the key relationships between firms' personnel strategies and worker and firm outcomes. The first three chapters in this volume give an excellent introduction into three of the most prominent and promising directions in this growing research area. Each of the three chapters is based on a unique new data source and addresses a core relationship between human resource practices at a different level of aggregation. The first chapter uses a large matched administrative employer-employee data set from the United States to analyze the relationship between human resource practices, research and development (R&D), and worker productivity in a sample of firms in the electronics industry. The second chapter augments a

2. See, for example, Abowd, Haltiwanger, and Lane (2004) and Brown, Haltiwanger, and Lane (2006).

traditional analysis of a large firm's personnel records with innovative survey data on workers' preferences, attitudes, and behaviors in a social context to study the determinants of turnover and worker productivity in the trucking industry. The third chapter tests predictions of a model of subjective evaluation and effort in a moral hazard setting using data from an exceptionally detailed matched worker firm survey from the French manufacturing industry.

Firms constantly face the problem of adjusting their production processes and their workforce to impulses from technological progress and increasing competition. A recurring theme is the question of how firms adjust their human resource (HR) practices to cope with the evolving economic environment, and whether some HR practices are more helpful in successfully implementing or developing new technologies than others. A core difficulty in providing an empirical answer to this question is that most data sets that contain information on investment into new technologies and R&D for a sufficient number of firms typically have little information on firms' HR structures. The first innovation of the first chapter, by Andersson, Brown, Campbell, Chiang, and Park, is to construct measures of HR practices based on longitudinal earnings and turnover information from the universe of workers in a large sample of firms in a particular industry (electronics). The chapter then analyzes the joint of occurrence of indicators such as accession and separation rates or within-job wage growth in HR "clusters" and describes the practices of firms with high and low investment in R&D. In a last step, the authors examine whether the interaction between R&D and HR practices significantly affects worker productivity. The chapter then interprets its tremendous amount of new information in the context of an economic model where firms have to decide whether to produce technology (R&D) in house or acquire it in the market and have to structure their HR practices to train, retain, or hire the appropriate workforce. The chapter's comprehensive descriptive empirical approach based on explicit firm-level HR measures grounded in economic intuition should pave the way to further fruitful analysis of the incidence and effect of HR practices using increasingly available matched administrative employer-employee data.

One of the great benefits of this approach is the potential to analyze the personnel choices and their correlation with worker and firm outcomes for a broad range of firms, workers, and phenomena of interest. The price to pay for this gain in insight is the focus on broader measures of HR practices. While ideal for describing recurring patterns and correlations, sometimes the relationships between firms' internal institutions and workers' incentives and productivity emphasized by the theory are more subtle. This is especially true for more recent modeling approaches emphasizing behavioral aspects of the firm-worker relationship, such as trust or loyalty. The second chapter in this volume, by Burks, Carpenter, Götte, Monaco,

Porter, and Rustichini, describes an ambitious data collection effort and research agenda aimed at uncovering specific and intricate links between HR practices, worker incentives and attitudes, and worker turnover and productivity. To do so, the chapter combines longitudinal personnel records of a large trucking firm in the United States with a panel survey of a cohort of newly hired workers geared to elicit a detailed range of behavioral and preference parameters. Thereby, a key innovation is to obtain information on risk and loss aversion or cooperation through standard survey questions as well as responses to small laboratory experiments such as prisoners' dilemma games. The second chapter describes this tremendous effort in detail and puts it into the context of a statistical analysis of how turnover and effort evolve with tenure at the trucking firm. The results show a large and increasing amount of variance in productivity of truck drivers and a differential effect of selective exit on variance as job tenure increases. While traditional analyses of firms' personnel records typically have to stop at this point, the added survey data will enable the authors to draw more specific conclusions about the determinants of turnover and the role of behavioral factors in future work. The strategy of combining firms' personnel records with innovative and detailed survey information documented in this chapter indicates another potentially highly fruitful area for future work.

While the second chapter exploits detailed information on a single firm to gain insights into particular aspects of the structure and effect of HR practices, sometimes representative surveys also contain information on both workers and firms amenable to a study of certain personnel policies. Such is the case with the French survey of Computerization and Organizational Change that collects information on team production and evaluation strategies for a small sample of workers in a representative sample of French firms. The third chapter, by Diaye, Greenan, and Urdanivia, exploits this source of information to test a model of the effect of subjective evaluation of workers' effort via interviews in the context of team and group work. Although increasingly common in practice, the analysis of evaluation interviews is rendered difficult due to complex interactions of various incentives, a lack of appropriate data, and identification issues due to unobserved heterogeneity. The chapter extends existing theoretical work to derive various predictions of how evaluation interviews should affect workers' effort and pay and uses propensity score methods to test this prediction, controlling for selection. The results indicate that evaluation interviews both attract high-productivity workers and have a direct effect on productivity within worker type. As survey data sets incorporate more detailed measures on work effort, organization, and HR practices, similarly ingenious combinations of specific theoretical modeling and statistical methodologies aimed at identification in other areas of personnel economics should become more common.

Firm Differences in Human Resource Practices

The chapters in this section provide evidence on differences in human resource practices by firms from a variety of different perspectives: over time (von Wachter and Bender), across firms (Hellerstein, Neumark, and McInerney), and within a firm (Manchester).

How much firms affect wages has been at the core of the literature analyzing the interrelationship between firms and workers. The evidence is clear that such firm-specific factors as firm size, unionization, and industry structure have important impacts. The interesting contribution of the von Wachter and Bender chapter is to show for a broad sample of establishments that some firm-specific effects are a function of the initial entry-level conditions that exist when a cohort of workers is hired. As the authors point out, there are two possible reasons for these entry-level differences. The first of these is that the quality of jobs and career opportunities differ for two cohorts; the second is that the degree of rent sharing between workers and firms depends on the extant market conditions at the time of entry. Their analysis exploits their longitudinal information on individuals, together with the firms that hire them, to show that at least some of the substantial wage differences that are observed across firms are due to transitory rents and disappear over time.

Another strand of the literature has focused on the role of labor market segregation in explaining race and sex wage differentials. A related strand has focused on workplace segregation by skill, as the productivity of more-educated workers has increased relative to less-educated ones. However, much of the work has focused on explaining cross-sectional differences across establishments, with the inevitable resulting concerns about omitted confounding factors. The work by Hellerstein, Neumark, and McInerney provides extremely useful initial evidence about both levels of segregation in the United States in 1990 and 2000 as well as changes. They find that racial and ethnic segregation is pervasive. And, while observed segregation by ethnicity has changed little, segregation by race has increased substantially. Most interestingly, the longitudinal nature of their data make it possible for them to show that the increase in racial segregation has been exacerbated by the entry and exit of establishments and by the changing industrial composition of the United States. Hellerstein, Neumark, and McInerney find that segregation by sex, even after controlling for occupational differences and despite countervailing industrial changes, has declined. Segregation by education is also substantial and has increased slightly over the decade.

The third chapter advances our understanding about a third HR practice in which substantial firm heterogeneity has been observed: training. Becker's seminal work suggests that firms will only provide specific train-

ing and that workers will bear the full cost of general training as general training will increase the likelihood that workers will leave the firm. Despite this theoretical prediction, there is not only abundant evidence that many firms offer tuition reimbursement programs, but also evidence that firms do this to reduce turnover. The Manchester chapter uses primarily a case study approach to examine whether there is an empirical basis that supports the notion that turnover is reduced. She finds that the five-year separation probability of workers who get tuition reimbursement is reduced by over 50 percent, but suggests that the reason is that the investment is complementary to firm-specific human capital. Hence, her results reconcile empirical observation with the Becker theoretical prediction.

Effects of Ownership Changes on the Organization of Production

While the previous sections featured chapters that examined differences in the ways in which firms treated workers, this section turns to examining what happens within firms as a result of changes in such fundamental features as ownership structure. The transition experienced by economies of Central and Eastern Europe provides a unique opportunity to examine such effects. The two studies on the effects of ownership in this volume, while quite different in their approach to the subject, both represent work that advance our understanding of one such key aspect of transition, the effects of ownership on wages and HR policies. The chapter by Earle and Telegdy uses a large linked employer-employee database for Hungary to analyze the effects of state and foreign ownership on wages. The chapter by Friebel and Panova is a case study of the HR policies in one insider-privatized Russian firm.

A key advantage of the Earle and Telegdy chapter, relative to previous studies of the effects of ownership that are based on either firm-level or worker-level data, is the use of longitudinal employer-employee data, which allows for identification of ownership effects taking into account differences in worker characteristics as well as nonrandom selection of firms into ownership status. The raw estimates show large wage differences across ownership types, but the authors find that ownership type is highly correlated with the education, experience, gender, and occupation of workers, suggesting ownership type may be systematically selected with respect to such characteristics. Nonetheless, the large unconditional wage gaps in the data are little affected by conditioning on worker characteristics. Ownership type is also correlated with firm size, industry, and productivity; controlling for industry reduces the estimated gaps, and controlling for employment size reduces them further. The chapter also exploits the presence of many switches of ownership type in the data to estimate firm fixed effects and random trend models, accounting for unobserved firm characteristics affecting the average level and trend growth of wages. The results

from these specification differ little in their implications for the effect of privatization, but they reduce the estimated effect of foreign ownership. Overall, the results imply that the substantial unconditional wage differentials are mostly, but not entirely, a function of differences in worker and firm characteristics and that linked panel data are necessary to take these correlated factors into account.

Why the operations of firms governed by private owners are quite different from those that are foreign-owned is something that has received a great deal of attention in the literature. While the Hungarian privatization process was not insider-dominated in the same way as other East European economies, most notably Russia, one hypothesis is that insider privatization results in little behavioral change. The fact that insiders benefited much from privatization in Russia raises suspicion about the efficiency of some of the privatization policies, and results suggest that insider-privatized firms do not restructure. However, the study by Friebe and Panova sheds new light on this puzzle in the transition literature by drilling down into the HR practices and the internal labor market of a single insider-privatized Russian firm. The results show career paths prior to transition that are quite similar to the career paths in western firms. In contrast to previous beliefs they find strong micro-evidence for restructuring activities after insider privatization. The employment of blue-collar workers decreased substantially, white-collar workers are recruited from outside the firm, while incumbent white-collar workers are shifted across functions within the firm (but do not leave the firm). As a result, the firm becomes “top loaded,” and career paths from lower levels in the hierarchy are effectively blocked, which consistent with the internal labor market literature can have adverse effects on the efficiency of the firm (see Gibbons and Waldman 1999).

In summary, privatization of state-operated enterprises is a key aspect of transition. Both these chapters present evidence that firm governance affects the HR practices and efficiency of firms, but perhaps in more complicated ways than economists thought when first tasked with forming transition policies in Central and Eastern European countries. Indeed, these two chapters highlight the relative unpreparedness of the economics profession to fully understand the impact of transition policies.

Globalization, Trade, and Labor Markets

Economists have long theorized about the impact of globalization and trade on the earning and employment outcomes of workers. Unfortunately, because the impacts of trade are typically measured at the firm level, and the policy interest is on the long-run outcomes of workers, little data have been available to examine the impact. These three chapters provide some of the first empirical evidence on the topic in analytic work only possible because of the existence of linked longitudinal employer-employee

(LEE) data.³ The chapters also highlight the new availability of LEE data, which were only available in Northern Europe a decade ago. Muendler uses Brazilian data, the Becker and Muendler chapter examines German multinationals, and Van Biesebroeck uses data from three African nations.

The very thought provoking piece by Muendler sets out to examine an issue of concern in every country: the impact of trade liberalization on workers. While economists at the International Monetary Fund (IMF) and the World Bank have long advocated liberalization as a path to economic growth and prosperity, arguing that the resulting reallocation of resources and economic growth would result in the absorption of displaced workers in the growing part of the economy. The rise of socialist leaders throughout South and Central America attest to the lack of popular confidence in such economic theories. Brazil's experience with trade liberalization in 1990 provides a useful opportunity to examine the facts. Muendler uses LEE data to examine the long-run outcomes of individual workers who worked in firms directly subject to foreign competition (i.e., in sectors in which foreign import penetration increased substantially) and compared them to observationally equivalent individuals in observationally equivalent firms. He finds, as expected, substantial displacement of workers; they also find that neither comparative-advantage sectors nor exporters absorb displaced workers for years. In addition, firms in the new-growth sectors have significantly more displacements and significantly fewer accessions than the exiting firms in the import-competing sector. As a result, workers are much more likely to transition to the informal sector and unemployment. Spells of unemployment last longer, and spells in the formal sector are much more likely to fail.

Another hotly debated impact of globalization is the outsourcing of jobs by multinational enterprises (MNEs). Lou Dobbs, a CNN reporter wrote a recent book, *Exporting America*, which has received enormous attention in the United States. He vehemently argues that too many U.S. companies are sending American jobs overseas and choosing to employ cheap overseas labor—going so far as to list “job exporters” on his Web site. A very different picture is painted in the Becker and Muendler chapter that uses German LEE data to examine the facts in some detail. They find that MNEs that increase their foreign direct investment (FDI) exposure become more competitive, and the resulting expansion acts to significantly reduce the rate of job loss. Indeed, the annual separation rate of workers at MNEs is about 14 percent, compared with the 18 percent separation rate of non-MNEs. One important result that is inconsistent with prior expectations

3. Although there has been some work on this topic by Lori Kletzer (<http://econ.ucsc.edu/Faculty/facLkletz.shtml>) or by Chris Ruhm (<http://www.nber.org/papers/w5621>), the studies are essentially based on worker surveys. Muendler is the first to use LEE data to analyse the long-term effects in a developing country.

about the nature of Germany's comparative advantage is that job savings are greater for more-educated than less-educated workers.

Although the previous two chapters deal with employment outcomes, policymakers and economists are equally interested in the impact of globalization on earnings. This is particularly interesting in the case of Africa, which has been receiving increasing attention in the development literature, but is probably underresearched due to lack of adequate data. The Van Biesebroeck chapter examines patterns in earnings outcomes of workers in manufacturing plants in three countries that differ substantially in levels of economic development: Zimbabwe, Kenya, and Tanzania. He finds that the more developed a country is (and the more exposed to foreign markets), the more wages match the productivity of the individual worker, controlling for other characteristics. Intriguingly, given the importance placed on education as an investment strategy for developing countries, wages substantially exceed productivity for the most-educated workers in the least-developed countries. This research provides some intriguing evidence that suggests that more-developed countries are also likely to have more efficient labor markets—although obviously the direction of causality is difficult to establish.

Data Sources

The data sets used in this book are truly international in flavor—and in a sense provide a world tour of the interrelationships between firms and workers. Four chapters (Andersson, Brown, Campbell, Chiang, and Park; Hellerstein, Neumark, and McInerney; Burks, Carpenter, Götte, Monaco, Porter, and Rustichini; and Manchester) use data on U.S. firms and workers. One of the first analyses of South American data is provided by the Muendler chapter, which studies Brazil. Crossing the ocean to Africa, Van Biesebroeck's chapter provides a study using data from Tanzania, Kenya, and Zimbabwe. The world tour continues with a trip north to the western part of Europe, with chapters from France (Diaye, Greenan, and Urdanivia) and Germany (Becker and Muendler; Bender and von Wachter). It ends in Eastern Europe (Russia, Friebe, and Panova) and Hungary (Earle and Telegdy).

One of the themes of the conference was to see how qualitative data analysis could be used to examine employers and employees. There are three case studies in the book; two are based on administrative records of the firm (Manchester; Friebe and Panova). Burks, Carpenter, Götte, Monaco, Porter, and Rustichini is a statistical case study of a single trucker firm and its employees, which matches proprietary personnel and operational data. They combine traditional survey instruments with behavioral economics experiments.

The other chapters contribute to the second theme of the book, which

was quantitative data, primarily LEE data. Although each chapter has the LEE component in common, a major contribution of the book is the description that the authors provide of how they create their LEE data sets by combining different sources with different identification techniques. The basic approach is well illustrated by von Wachter and Bender, who use a single data source for building up their linked employer employee data. And, although most of the chapters use different official data sets and are linked over a unique firm identifier, two chapters (Becker and Muendler; Hellerstein, Neumark, and McInerney) illustrate the use and validity of alphanumeric matching algorithms, such as Automatch, based on the name and addresses of the firms. A different approach is used by Earle and Telegdy who have neither identifiers nor addresses as a basis for linking one data set: they use subsets of common variables to get unique firms per cell for matching.

Other striking features are evident upon examination of the chapters. The first is the imaginative set of methods whereby data can be collected from individual firms; the second is how many more data sets have become available since the seminal review by Abowd and Kramarz in 1999; and the third is the number of innovative approaches taken to enhance the breadth and depth of information derived from linked data.

Several chapters illustrated the potential richness of looking at single firms in detail. One example is the chapter by Manchester, who analyses the impact of tuition reimbursement programs by examining data from a single firm as well as from a cross section of firms. A panel of observations was constructed based on seven point-in-time observations from administrative records for the firm. Individuals are observed on December 15 of each year from 1999 to 2005, and the data include gender, age, and race as well as start date, job characteristics, and annual wage rates. The chapter by Friebel and Panova also uses the raw data from the HR department of the firm they study. There is one personnel file for each individual in the firm. This file contains information about entry into the firm; exit date (that is, separation); dates of movements across job titles; an occupational code; the departmental affiliation; as well as such personal characteristics as age, education, gender, place of birth, place of university education, field of study, party and trade union membership, ethnicity, marital status, and number of children. Finally, the firm provided them with information about the job history of each individual: military service, date of leaving previous job, last employer. In probably the richest case study analysis in the book, the chapter by Burks, Carpenter, Götte, Monaco, Porter, and Rustichini show the enormous potential of a detailed data collection effort. The team matched proprietary personnel and operational data to new data collected by the researchers to create a two-year panel study of a large subset of new hires that included both standard survey instruments and the results of behavioral economics experiments. The team collected information on employees that

went beyond the standard demographic information to also capture information on risk and loss aversion, time preference, planning, nonverbal IQ, and the Multidimensional Personality Questionnaire (MPQ) personality profile. The project, which is in its beginning stages, will follow employees over two years of their work lives. Among the major design goals are to discover the extent to which the survey and experimental measures are correlated and whether and how much predictive power, with respect to key on-the-job outcome variables, is added by the behavioral measures.

Table I.1 provides a brief summary of the LEE data sets described in each chapter, together with a reference to the Abowd and Kramarz (1999) paper. As is to be expected, data sets that already existed have been considerably updated and enhanced; in a number of cases, new data sets have appeared.

The basic richness of LEE data is well illustrated by the von Wachter and Bender chapter: the size and universality of the data mean that they can track the outcomes of quite narrowly defined subgroups of workers as well as the firms that employ them. Their data includes information on all employees covered by social security, representing around 80 percent of the German workforce, with detailed histories for each worker's time in covered employment. Their data are unusually rich for LEE data as they not only include basic demographic information, as well as data on occupation, industry, job status, education, and individual-level information on gross daily wages subject to social security contributions and the exact dates when the employee worked in the social security system. The unique establishment identifiers available were used to create a separate data set of establishment characteristics that were aggregated up from the employment register and merged back onto the individual-level data. Characteristics include, among others, establishment size, employment growth, and average wages.

Muendler's chapter features one of the first LEE data sets for a developing country: Brazil. The data set is derived from a nationwide, comprehensive set of administrative records of workers employed in the formal sector. The ministry of labor estimates that well above 90 percent of all formally employed workers in Brazil are covered in *Relação Anual de Informações Sociais* (RAIS) throughout the 1990s. As is typical with LEE data derived from administrative records, the sample size is enormous: information on 71.1 million workers (with 556.3 million job spells) at 5.52 million establishments in 3.75 million firms over the sixteen-year period 1986 to 2001. The data also have the rich feature of providing the month of accession and the month of separation in addition to such other relevant worker information includes tenure at the establishment, age, gender, and educational attainment; job information includes occupation and the monthly average wage; establishment information includes sector and municipality classifications. As is also often the case with LEE data derived from administra-

Table I.1 Summary of linked employer-employee data sets

Author	Country	Brief description	Abowd & Kramarz comparison
Andersson et al.	United States	a. Longitudinal individual data from the LEHD b. 1997 Economic Censuses c. The 1991–1998 Census/NSF R&D Surveys (firm-level R&D)	New data set
Diaye et al.	France	COI survey	Enhanced data set
Von Wachter & Bender	Germany	Social security records—total population	Referenced in Abowd & Kramarz
Hellerstein et al.	United States	1990 and 2000 DEEDs	Updated and expanded
Earle & Telegdy	Hungary	a. Hungarian Wage Survey b. Tax Authority data	New data set, new country
Muendler	Brazil	RAIS	New data, new country
Becker & Muendler	Germany	a. Social security records of the German Federal Labor Agency b. Midi database (MIcro database Direct Investment, formerly direk), collected by Deutsche Bundesbank c. Commercial corporate structure database markus (from Verband der Vereine Creditreform)	Updated and enhanced (by b and c)
Van Biesebroeck	Tanzania, Kenya, and Zimbabwe	A stratified sample of manufacturing firms in three consecutive years	New data sets, new countries

tive records, the basic data set can be enhanced by matching into other administrative records—in this case, export data.

The Andersson, Brown, Campbell, Chiang, and Park chapter further illustrates the potential for enhancing basic LEE data by using multiple data sources. In order to answer their research question, they use data from three sources: longitudinal and near-universal individual data from the Longitudinal Employer-Household Dynamics Program (LEHD) program to construct and characterize the HR practices of firms; they add firm characteristics from the 1997 Economic Censuses (e.g., measures of revenue, material costs, total hours, capital stock, four-digit industry code) as well as from the 1991 to 1998 Census National Science Foundation (NSF) R&D Surveys (firm-level R&D).

A very similar approach is taken by Becker and Muendler, who construct their LEE set from three confidential micro-data sources, from multiple sources, and complement them with sector-level information on

German foreign trade. This chapter illustrates some of the matching challenges common to LEE data, in that the three data sources do not share common firm identifiers. The authors surmount the problem by using a string-matching procedure to identify clearly identical firms and their establishments. The result of their efforts is to create an impressively large data set that constitutes a cross section of establishments around year 2000, including a total of 39,681 establishments whose German parent firms conduct FDI abroad and 1,133,920 control establishments out of 3.8 million establishments in the full worker sample (1998 to 2002).

Not all LEE data are derived from official administrative records. Indeed, Van Biesebroeck's uses stratified samples of manufacturing firms in three African countries that were collected by three different research teams, coordinated by the Regional Program of Enterprise Development at the World Bank. When working with survey data, one of the most difficult decisions is whether to choose a sample that is representative of firms or one that is representative of workers. In this case, the latter was chosen: firms were sampled to give (the firm of) each manufacturing worker equal probability to be included in the sample—an implicit stratification by employment size. The second choice is how many employees to interview within each firm. The cost constraints that are an integral part of survey work impose a trade-off between firm and worker diversity: the more employees interviewed in a firm (and, hence, the more within firm diversity is captured), the fewer firms can be interviewed (resulting in less across-firm diversity). In the African countries, a maximum of ten employees per firm were interviewed each year, resulting in an unbalanced panel of firms with, on average, 110 to 183 observations per year in each country.

Another example of a survey approach is evident in the Diaye, Greenan, and Urdanivia chapter, which analyzes a French matched employer-employee survey on computerization and organizational change. In contrast to the African case, however, the sampling unit is the firm, and the frame is a representative sample of manufacturing firms with more than fifty employees and a sample of randomly selected employees within these firms. The French sample chose the opposite trade-off from the African, interviewing a small sample of employees (one, two, or three) within each firm and, hence, getting more firm diversity. An interesting feature of the French survey, and one that emphasizes the value added of surveys relative to administrative records, is that the labor force section provides a detailed description about the organization of work, particularly whether that work is structured around group activities. In addition, the survey captures different measures of effort, which would be impossible to capture in administrative records.

The Earle and Telegdy chapter also draws on data from multiple sources. The worker data come from the Hungarian Wage Survey, compiled by the National Employment Office and maintained by the Institute of Econom-

ics of the Hungarian Academy of Sciences. The latter organization also links these data to some firm-level information. The authors have further linked this database (using common variables in both databases) to a universal tax database containing detailed information on all Hungarian firms using double-sided accounting. Employers are included in the Wage Survey according to whether their employees are selected by a random procedure: in the first two years of the survey, workers are selected using a fixed interval of selection, while subsequently workers are selected by birthdate. Firms were included only if they have employees born on these dates. Although this approach provides a random sample of workers within firms and includes, on average, about 6.5 percent of production workers and 10 percent of nonproduction workers, the sample of firms is related to size. The authors, therefore, use weights related to size and response probabilities in their analysis, and the final sample consists of a panel of 21,238 firms linked with a within-firm random sample of 1.35 million worker observations.

The final approach that can be used to create LEE data sets is to match existing surveys with existing administrative data. This approach is described in the Hellerstein, Neumark, and McInerney chapter. They draw a sample of workers from the Sample Edited Detail File (SEDF), which contains all individual responses to the Decennial Census of Population one-in-six Long Form. Respondents are asked to provide the name and address of their employer in the previous week. This information is then matched to the name and address information in the Census Bureau's Business Register list (BR), based on administrative records. This procedure yields a very large data set with workers matched to their establishments, along with all of the information on workers from the SEDF. Even more interesting, because the data are constructed for both 1990 and 2000, the authors are able to examine changes in establishments over time.

Summary and Outlook

In summary, the chapters in this volume all represent research that relies on advances in data collection methods in one way or another. These range from combining case study data with personnel records of a single firm, ideally suited to understand issues such as how HR policies affect workers and the performance of firms, to the creation of new multi-source, nationally representative LEE data sets, ideally suited to capture empirical irregularities related to the dynamics of the economy. It is easy to envision further advances along each of these lines, but perhaps even more promising is the crossing of the two paths. The studies in this volume clearly demonstrate the relative advantages of the two approaches and the results often complement each other in a way that adds to our overall understanding.

The benefits of being able to combine data from the two study approaches within a single framework are obvious, as is the increasing po-

tential for cross-national comparisons (see Hamermesh 2007). Today they are more feasible than before thanks to new data collection tools and new integration techniques. Indeed, active research is in progress to match data sets from many different sources to inform research about active labor market policies, price setting, and employment dynamics. Similarly, international comparisons of LEE and firm surveys are beginning to emerge (Lazear and Shaw, forthcoming; Freeman, Kruse, and Blasi, forthcoming).

However, there are clearly important and big challenges in terms of data access issues and disclosure avoidance that need to be addressed. The creation and analysis of high-quality information are core elements of the scientific endeavor. No less fundamental is the ability to replicate scientific analysis. Yet the firm-level data that is described in this book are often not accessible to others for replication and validation. It will be critical to develop widely available access modalities for the qualitative data, often housed at universities and research centers, and administrative data, often housed at statistical agencies. Only recently has the research community begun to address such key issues that will ultimately affect the scientific nature of the research as well as our ability to access and gather new data. Some progress has already been made, including the development of remote access capabilities and new synthetic data methods, but this is likely to remain a key challenge for the research community in the foreseeable future.

References

- Abowd, John, John Haltiwanger, and Julia Lane. 2004. Integrated longitudinal employer-employee data for the United States. *American Economic Review* 94:224–29.
- Abowd, John, and Francis Kramarz. 1999. The analysis of labor markets using matched employer-employee data. In *Handbook of labor economics*. Vol. 3B, ed. O. Ashenfelter and D. Card, 2629–2710. Amsterdam: North-Holland.
- Brown, Clair, John Haltiwanger, and Julia Lane. 2006. *Economic turbulence: Is volatility good for America?* Chicago: University of Chicago Press.
- Freeman, Richard B., Douglas Kruse, and Joseph Blasi. Forthcoming. Worker reports on labour practices and outcomes in a single form across countries. *Labour Economics*.
- Gibbons, Robert, and Michael Waldman. 1999. A theory of wage and promotion dynamics in inside firms. *Quarterly Journal of Economics* 114 (1999): 132–58.
- Hamermesh, Dan. 2007. Fun with matched employer-employee data. *Labour Economics*, forthcoming.
- Lazear, Edward P., and Kathryn L. Shaw, eds. Forthcoming. *An international comparison of the structure of wages*. Chicago: University of Chicago Press.