

WORKING P A P E R

A Description and Analysis of Evolving Data Resources on Small Business

AMELIA HAVILAND AND
BOGDAN SAVYCH

WR-293-1-ICJ

September 2007

This product is part of the RAND Institute for Civil Justice working paper series. RAND working papers are intended to share researchers' latest findings and to solicit additional peer review. This paper has been peer reviewed but not edited. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.



KAUFFMAN-RAND INSTITUTE FOR
ENTREPRENEURSHIP PUBLIC POLICY

PREFACE

Data quality is an important determinant of the quality of research on small business and entrepreneurship. Historically, data collected on U.S. businesses have focused almost exclusively on large firms (typically those with at least 250 employees); the number, quality, and richness of data sources on small firms, while improving, remain limited. This document describes the main government and private data sources currently available or under construction for research on small business and entrepreneurship. The paper also provides a listing of resources researchers can use to gain more information about each data set.

This information should be of interest to researchers, but also to policymakers at both the state and federal level who use research to inform their decisionmaking. The work was completed under the auspices of the Kauffman-RAND Institute for Entrepreneurship Public Policy and was funded by the Ewing Marion Kauffman Foundation.

The RAND Institute for Civil Justice is an independent research program within the RAND Corporation. The mission of the RAND Institute for Civil Justice (ICJ), a division of the RAND Corporation, is to improve private and public decisionmaking on civil legal issues by supplying policymakers and the public with the results of objective, empirically based, analytic research. The ICJ facilitates change in the civil justice system by analyzing trends and outcomes, identifying and evaluating policy options, and bringing together representatives of different interests to debate alternative solutions to policy problems. The Institute builds on a long tradition of RAND research characterized by an interdisciplinary, empirical approach to public policy issues and rigorous standards of quality, objectivity, and independence.

ICJ research is supported by pooled grants from corporations, trade and professional associations, and individuals; by government grants and contracts; and by private foundations. The Institute disseminates its work widely to the legal, business, and research communities, and to the general public. In accordance with RAND policy, all Institute research products are subject to peer review before publication. ICJ publications do not necessarily reflect the opinions or policies of the research sponsors or of the ICJ Board of Overseers.

The Kauffman-RAND Institute for Entrepreneurship Public Policy, which is housed within the RAND Institute for Civil Justice, is dedicated to assessing and improving legal and regulatory policymaking as it relates to small businesses and entrepreneurship in a wide range of settings, including corporate governance, employment law, consumer law, securities regulation and business ethics. The institute's work is supported by a grant from the Ewing Marion Kauffman Foundation.

For additional information on the RAND Institute for Civil Justice or the Kauffman-RAND Institute for Entrepreneurship Public Policy, please contact:

Robert Reville, Director
RAND Institute for Civil Justice
1776 Main Street, P.O. Box 2138
Santa Monica, CA 90407-2138
(310) 393-0411 x6786;
FAX: (310) 451-6979
Email: Robert_Reville@rand.org

Susan Gates, Director
Kauffman-RAND Institute for Entrepreneurship
Public Policy
1776 Main Street, P.O. Box 2138
Santa Monica, CA 90407-2138
(310) 393-0411 x7452;
FAX: (310) 451-6979
Email: Susan_Gates@rand.org

ACKNOWLEDGMENTS

This research was conducted with the Kauffman-RAND Institute for Entrepreneurship Public Policy. The Center is funded by a grant from the Kauffman Foundation to the RAND Corporation for research dealing with regulatory issues and small businesses.

We are grateful for the input we received from a number of sources in creating this document. We received substantial and helpful suggestions on early drafts of the manuscript from Susan Gates. The thoughtful and constructive comments from reviewers Pinar Karaca-Mandic and Jacob Klerman improved the manuscript considerably, and we thank them for their time and attention. Kristin Leuschner helped to improve the presentation and flow of the document. Lisa Bernard, Michelle Platt and Terry Lane helped with the editing and preparation of the manuscript. Any remaining errors are, of course, the sole responsibility of the authors.

ABBREVIATIONS

BED	Business Employment Dynamics
BITS	Business Information Tracking System
BLS	U.S. Bureau of Labor Statistics
BR	Business Register
CBO	Characteristics of Business Owners
CBP	County Business Patterns
CES	Current Employment Statistics
CEW	Covered Employment and Wages
CFI	Census of Finance, Insurance and Real Estate
CFN	Census File Number
CMF	Census of Manufacturers
COS	Company Organization Survey
CPS	Current Population Survey
CRSP	Center for Research in Security Prices
CRT	Census of Retail Trade
CSR	Census of Services
CUT	Census of Transportation, Communications, and Utilities
CWH	Census of Wholesale Trade
D&B	Dun and Bradstreet
DMI	Dun's Market Identifier
EBS	Employee Benefit Survey
EC	Economic Census
ECI	Employment Cost Index
EIN	employer identification number
HC	Household Component
IC	Insurance Component
ICJ	Institute for Civil Justice
ILBD	Integrated Longitudinal Business Database
Kaiser/HRET	Kaiser Family Foundation/Health Research and Educational Trust
KFS	Kauffman Firm Survey
LBD	Longitudinal Database of Establishments Covered by State Unemployment Insurance Programs
LEEM	Longitudinal Enterprise and Establishment Microdata

LEHD	Longitudinal Employer Household Dynamics
MEPS	Medical Expenditures Panel Survey
MH	Martindale-Hubbell Law Directory
MPC	Medical Provider Component
MSA	metropolitan statistical area
NAICS	North American Industry Classification System
NCHS	National Center for Health Statistics
NCS	National Compensation Survey
NEHIS	National Employer Health Insurance Survey
NHC	Nursing Home Component
OCS	Occupational Compensation Survey
PPN	Permanent Plant Number
QCEW	Quarterly Census of Employment and Wages
RDC	Research Data Center
SBA	U.S. Small Business Administration
SBO	Survey of Business Owners
SESA	State Employment Security Agency
SESA-ID	State Employment Security Agency identification number
SIC	standard industrial classification
SIPP	Survey of Income and Program Participation
SMOBE	Survey of Minority-Owned Business Enterprises
SSA	Social Security Administration
SSBF	Survey of Small Business Finances
SSEL	Standard Statistical Establishment Listing
SSN	social security number
SUSB	Statistics of U.S. Businesses
SWOBE	Survey of Women-Owned Business Enterprises
UCFE	Unemployment Compensation for Federal Employees
UI	unemployment insurance
ZBP	ZIP Code Business Patterns

INTRODUCTION

Historically, data collected on U.S. businesses have focused almost exclusively on large firms (typically those with at least 250 employees). As a result, researchers interested in small businesses and entrepreneurship have been strongly constrained in their ability to carry out empirical, policy-related research. Ongoing concerns about the lack and quality of data on small firms led to a 2004 conference on data sources related to entrepreneurship,¹ the creation of a 2004–2006 National Academy of Sciences panel on Federal Business Statistics, and a 2006 National Research Council Workshop on improving business statistics through interagency data sharing (Kuebler and Mackie, 2006). Such efforts have been part of a more recent trend to increase the number, quality, and richness of data sources on small firms. While data sources continue to improve, information on the uses, availability, and limitations of these sources is scattered among the multitude of governmental and private organizations that collect and own the data.

Policy-relevant research on small business is typically concerned with questions of how a particular policy has or will influence small firms. It is thus concerned with changes in outcomes for small businesses over time as well as comparisons between small businesses and larger businesses. As a result, data relevant for small business research are typically data that are relevant for research on business more generally, with the added requirement of the availability of information on the size of the business.

Empirical research on small business and entrepreneurship must begin with a definition of what constitutes a small business. The simplest definition, and the one most frequently used by the U.S. Small Business Administration (SBA), is a firm with fewer than 500 employees.² However, policymakers at the federal, state, and local levels use a wide variety of approaches to define a small business either to include businesses in special support programs or to exempt them from specific regulations. These definitions may be based on a variety of company characteristics and often depend on the specific nature of the policy or regulatory issue under consideration. For example, for the purposes of employment and health insurance regulations, size is typically defined in terms of the number of employees in the firm, whereas economic regulation often classifies small businesses based on market capitalization or organizational form.

Once the proper definition of small business has been identified, researchers must assess the extent to which available data allow for the measurement of firm size and other outcomes of interest in a way that is faithful to that definition. If a necessary data set does not yet exist, researchers may instead need to carry out their own data collection using an appropriate sampling frame of firms or establishments. Ideally, this

¹ Kauffman Symposium on Entrepreneurship data (November 10–11, 2004).

² The definition of what constitutes a small business differs among industries. The Small Business Administration provides a complete list of small business size standards matched to the North American Industry Classification System (NAICS). The size standard is expressed in average annual receipts, average number of employees, or, sometimes, in physical volume of output. For more information see U.S. Small Business Administration (undated).

sampling frame would provide enough information to identify units of interest without screening. In assessing the value of a particular data set as a frame or for analysis, researchers need to consider several issues: available firm size measure (establishment or firm), source of data (employee or firm), access to data, data quality, and the ability to link data longitudinally. Here we briefly discuss why each of the issues listed are important for carrying out research on small businesses.

Available size measures: In addition to different measures of size, researchers may be interested in the sizes of different business units: establishments or firms. From a policy perspective, the relevant unit of analysis for small business issues is typically the firm. However, many existing data sets collect data at the level of establishment or work site. An establishment is an economic unit that produces goods or provides services, such as a factory, store, or mine. Usually, it is a single physical location engaged in predominantly one economic activity (see U.S. Bureau of Labor Statistics, 1997). Although it is common for researchers to use establishment-level data to study small business issues, it is crucial to recognize the differences between firms and establishments. At one extreme, a firm may be comprised of a single establishment that is the same size as the firm. At the other extreme, a firm of, say, 1,000 employees may be comprised of 1,000 establishments with one employee each. As illustrated in the work of Mendeloff et al. (2006), firm size and establishment size can have different relationships to outcomes of interest. Researchers must take care to clearly understand the unit of analysis of data and how well matched it is to the definition of small business.

Source of data (employer or firm): Data on a workplace can be obtained either from the employees or from the employer or firm. Each source of information has advantages and disadvantages, and these will vary depending on the specific policy question under consideration.

Access to data: Researchers are concerned about the ease of access to data. If data are publicly available, is there a fee associated with this access? If data are not publicly available, how long and involved is the process for accessing these data? If the data set is not publicly available, researchers may use aggregated tables created from the data. Most agencies provide tables aggregated by such features as geography and industry.

Data quality: How reliable and accurate are the data? If data are based on a survey, what are the sampling frame and the response rate to the survey? If data are based on administrative records, how complete and reliable is the recording process?

Ability to link data longitudinally: Studies of small business are often concerned with questions of entry and exit, as well as firm growth, prompting the question: Did a particular policy force small firms out of the market or impede their growth relative to that of larger firms? To study these issues, it is crucial for researchers to be able to link data over time to identify when new firms enter the data, when firms leave the data, and to determine the growth trajectory of firms. A key consideration is therefore whether it is possible to link data longitudinally, and if so, for how long.

This document briefly describes the main government and private data sources currently available or under construction that could be used for research on small business and entrepreneurship. The paper also provides a listing of resources researchers can use to gain more information about each data set.

For each data set, we describe the collection method, coverage, main variables and limitations, and data uses. We start the description of each data set with a discussion of how the data are obtained, whether from administrative records, survey, or census. Then, we identify the population from which the data are collected (e.g., firm, establishment, owner, or worker) and the unit of observation. Most of the business information is collected at the establishment level, a single location at which goods are produced or services are provided. In some cases, these establishment data can be aggregated to the level of firm, which might include several establishments. We also describe the main variables available in the data set and the main limitations of the data. We are especially interested in the variables that measure size of the firm because of the obvious relevance to small business research. We select employment as the main measure of firm size, although other measures such as revenues or sales are feasible and are available in some data sets. We conclude the description of each data set with a discussion of current and potential future uses of the data. Table A.1 in the appendix provides a summary reference concerning the data sets described in this report.

This document is organized into four main chapters. The Chapters Two through Four discuss government data sources: the U.S. Bureau of Labor Statistics (BLS), the Census Bureau, and other government sources. The Chapter Five describes private data sources. These chapters are followed by a brief discussion of ongoing data collection needs and a lightly annotated bibliography of additional references on the data sources described here as well as recent research on small businesses that make use of these data sources.

Chapter Two

BUREAU OF LABOR STATISTICS DATA SOURCES

A multitude of government agencies collect information relevant to small business research. The following section focuses on several data sets created by the Bureau of Labor Statistics (BLS). BLS is the principal federal agency that collects data in the field of labor economics and statistics. The mission of BLS is to “collect, process, analyze, and disseminate essential statistical data to the American public, the U.S. Congress, other federal agencies, state and local governments, business, and labor” (see BLS, 2001). Its goal is to provide timely, consistent, and high-quality data on a range of issues including employment, wages, and workers’ benefits.

To achieve its mission, BLS maintains a number of different data sets. Those include the Quarterly Census of Employment and Wages (QCEW), Business Employment Dynamics (BED), Current Employment Statistics (CES), and the National Compensation Survey (NCS). The QCEW and BED data are closely related. The QCEW is a comprehensive set of data on establishment-level employment and payroll information derived from administrative records. The BED data link QCEW data at the establishment level over time to create longitudinal histories. Although the data in the QCEW and BED reflect establishment-level information, researchers have used the employer identification number (EIN) to aggregate information to the firm level. The CES and NCS data are based on establishment-level surveys on employment, employee characteristics, hours, benefits, earnings, and organizational characteristics conducted by BLS. These surveys supplement the employment and payroll information available through the QCEW and BED. Because they reflect a sample rather than the universe of establishments, there is no way to aggregate the data to the firm level. However, with full data access, it should be possible to link the records from these surveys to QCEW data and then control for firm size in examining the establishment-level data. A key limitation of all four of these data sets is that they are not publicly available. Researchers interested in using these data must submit an application to BLS, including a description of how the proposed research will benefit BLS, and restrictions may be imposed on the analytical outputs.³ These data can provide a cross-sectional and dynamic picture of local labor markets and employment patterns. The BED in particular can be used to study within-establishment changes over time and across location, job creation, job destruction, and changes in establishment and firm sizes over time.

BLS also cooperates with the U.S. Census Bureau to collect data for the Current Population Survey (CPS), a household-level survey. The survey gathers demographic and labor market information, as well as information on a variety of specific topics such as education, health insurance, and pensions. Because it is a household-level survey, it captures information on self-employed individuals; however, firm-level information

³ For the list of available data sets and application rules, see BLS (2007b).

is extremely limited. Households are surveyed several times over the course of two years, so it is possible to create a short longitudinal series for individuals. Among other things, this allows researchers to identify new business owners. The coverage of the CPS makes it a particularly useful tool for the analysis of the self-employed and new entrepreneurs. Unlike the other data sources described in this chapter, the CPS data are publicly available.

The remainder of this chapter provides detailed information on each of the data sets discussed above.

QUARTERLY CENSUS OF EMPLOYMENT AND WAGES

Data Collection Method and Coverage

The Quarterly Census of Employment and Wages (QCEW), also known as the Covered Employment and Wages (CEW) and ES-202 program, is BLS' most extensive effort to collect quarterly employment and payroll information. This comprehensive data set is the result of cooperation between BLS and the State Employment Security Agencies (SESAs). The SESAs derive data from quarterly contribution reports submitted by employers. All employers are required to pay quarterly taxes based on the employment and wages of workers covered by state unemployment insurance (UI) laws and federal workers covered by the Unemployment Compensation for Federal Employees (UCFE) program. The data are derived from the administrative records that employers need to submit each quarter to the SESAs for each of their establishments.

As such, the data include all establishments subject to state and federal unemployment insurance laws. In 2004, the data were collected from about 8.4 million establishments, a majority of which are small.⁴ As described in the *BLS Handbook of Methods* (BLS, 1997), QCEW provides a virtual census of nonagricultural employees and their wages and reports that about 47 percent of all workers in agricultural industries are also covered. The data cover approximately 98 percent of all employment;⁵ the major exclusions from UI coverage are self-employed workers, religious organizations, most agricultural workers on small farms, active duty military personnel, elected officials in most states, most employees of railroads, some domestic workers, unpaid family workers, most student workers at school, and employees of certain small nonprofit organizations (note that the regulations determining which nonprofits are required to have UI coverage vary by state and may change over time). The quarterly data are available starting in 1975, although different spans of years are available with the two different sets of industry classification codes: 1975 to 2000 using Standard Industrial Classification (SIC) codes; 2001 forward using North American Industry Classification System (NAICS) codes; and from 1990 to 2000 using reconstructed NAICS codes.

⁴ Out of 7.9 million establishments in 2003, about 60 percent of establishments had fewer than five workers, another 17 percent had five to nine workers, and another 18 percent of establishments had 10 to 50 employees. Overall, about 99.8 percent of establishments had fewer than 500 employees. See BLS (2006a).

⁵ See technical note to the most recent report (BLS, 2007a).

Main Variables

Firm and Establishment Size Variables. The QCEW data set includes information on establishments' employment each month of the prior quarter and quarterly wages. Monthly employment data represent the number of workers who worked during, or received pay for, the pay period including the 12th day of each month. As explained in BLS (1997), the employment measure includes all corporation officials, executives, supervisory personnel, clerical workers, wage earners, pieceworkers, and part-time workers. Workers are reported in the state and county of the physical location of their job. This measure of employment includes all workers on paid sick leave, paid holiday, paid vacation, and so forth, but excludes those on leave without pay for the entire payroll period (see BLS, 1997).

The QCEW also provides data on quarterly wages; however, the definitions of what constitutes wages differ among states. In most states, total wages include gross wages and salaries, bonuses, stock options, tips, and other gratuities, as well as the value of meals and lodging, where supplied. Some states also include in total wages employer contributions to certain deferred compensation plans, such as 401(k) plans.

All variables in the QCEW are collected and provided at the level of establishments. Employers who operate multiple establishments within a state submit a multiple work site report that provides detailed information on each of the establishments. Although the data are provided at the level of establishment, one can aggregate it to the level of firm using IRS-assigned EINs. See more details in the discussion of the Business Employment Dynamics (BED) database.

Other Variables. Also available in the data are location (full address) and industry classification.

Major Limitations

The primary disadvantage of this data set is lack of accessibility. The data set, as well as a number of other data sets constructed by BLS, are not publicly available but can be accessed only in restricted settings at the BLS offices and after an application process.⁶ BLS accepts applications for use of the data set four times a year, and any research must benefit BLS. Both the application and the application time are relatively short, with a five- to ten-page proposal due on April 15 for use of the data starting in the summer. In addition, there are restrictions on the type of analysis outputs from the QCEW that can be released. Currently, for example, access to the confidential data of the QCEW program is considered for research projects that do not involve tabular output.

Although the data set is not publicly accessible, extensive aggregate tables created from these sources are publicly available. The QCEW program provides tables for employment and wages aggregated by geography (nation, state, metropolitan statistical area [MSA], and county) and industry.⁷ For the first quarter

⁶ For the list of available data sets and application rules, see BLS (2007b). BLS refers to the QCEW and the BED as the Longitudinal Database of Establishments Covered by State Unemployment Insurance Programs.

⁷ Prior to 2001, the QCEW collected and reported data based on the SIC system. After that, it started collecting data based on the NAICS and reconstructed aggregated tables for periods covering 1990 to 2000 using the new classification system. Note, however, that the reconstructed data set is not free of errors, and researchers should be very careful when using the reconstructed data. Some researchers even decide not to use the reconstructed data, given some large inconsistencies in the records. For more information, see BLS (undated[c]).

of each year, BLS also produces these tabulations by establishment employment categories.⁸ These tabulations are released only at the national and state levels. Researchers using the publicly available aggregated tables should be aware that BLS is withholding some information to prevent disclosure of individual employers.

In addition, researchers using the data set need to be aware of differences among states with regard to who is covered and what data are collected. Also, the year-to-year consistency of the employment and wage data might be affected by periodic changes in state and federal UI laws. For example, since January 1, 2004, the Washington Employment Security Department no longer includes as covered wages an employee's income attributable to the transfer of shares of stock to the employee. The details of coverage and differences among states are provided in the description of the state UI legislation.

Furthermore, the QCEW data set is not designed as a time series. However, the QCEW establishment-level observations can be connected over time to construct a longitudinal database. These linkages are used to create the BED database described below.

What Can Be Done with the Data

The QCEW provides cross sections of administrative data for a limited number of variables covering nearly the entire universe of establishments. So far, there is little research using cross sections from the QCEW. Access to data is a major constraint that researchers face in using the full records. Assuming that a researcher has obtained access to the data, the records collected through the QCEW program can be used in a number of situations. These confidential records include address of the business, which can be used to compare regional policies or compare effects of the policies in the affected and unaffected regions, counties, ZIP codes, or other geographic areas. Modern mapping software may provide a way to aggregate available data to some geographic area as described in Konigsberg, Talan, and Clayton (2005). In addition, BLS links these cross-sectional records to produce a longitudinal database of establishments. Researchers can gain access to a longitudinal data set of establishments and firms through the BED program discussed below. The Census Bureau also draws upon these data to supplement various multilevel business databases. In addition, this database can be used as a sampling frame for further surveys and to produce denominators for other research, if access issues are resolved.

Some researchers may also take advantage of the aggregated tables that are publicly available. These tables can be used to examine effects of the state-level policies on the distribution of the establishments by establishment-size category. Unfortunately, these tables are not provided at the firm level, and there is no information about entrance or exit from the market.

⁸ The March employment level determines the size category of each establishment. The establishment employment size categories are less than five workers, five to nine workers, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, and 1,000 or more employees per establishment. Note that each establishment of a multiestablishment firm is tabulated separately into the appropriate size category, and the total employment level of the reporting multiestablishment company is not used in the size tabulation.

BUSINESS EMPLOYMENT DYNAMICS

Data Collection Method and Coverage

The Businesses Employment Dynamics (BED) program connects quarterly establishment records from the QCEW administrative records over time to produce the Longitudinal Database of Establishments Covered by State Unemployment Insurance Programs, also known as the BED data set.⁹ The data are complete from September 1992 to September 2003 and are being expanded on an ongoing basis. These longitudinal data are more frequently updated and available on a timelier basis than anything previously available.

Pivetz, Searson, and Spletzer (2001) describe how the BED longitudinal linkages are developed. Most of the establishments are matched from quarter to quarter using the unique SESA identification number (SESA-ID). This approach, however, might miss possible links because of changes in ownership, firm restructuring, or UI account restructuring. In this case, probability-based matching is used to link establishments with different SESA-IDs. The match is based upon comparisons such as the same name, address, and phone number. Third, an analyst examines unmatched records individually and makes a possible match. The resulting data set includes longitudinal histories with quarterly data of more than 6.4 million private-sector establishments.

Given that the data set is derived from the QCEW establishment records, one should not expect large differences in coverage and definition of variables between QCEW and BED. The major contrast existing between the QCEW data and the BED data set is that the latter excludes government employees, private households, and establishments with zero employment.¹⁰

Main Variables

Firm and Establishment Size Variables. Most of the variables in the BED are the same as in the QCEW. The data set provides monthly records of establishment employment and quarterly records of establishment wages. Using these data, one can track net employment changes at the establishment level over time and determine job gains and losses at expanding and contracting establishments.

More recent work has also presented BLS establishment-level data aggregated to the firm level using IRS-assigned EINs (see Okolie, 2004). Using the BED data, Okolie (2004) presented net employment changes at both the establishment and firm levels. The underlying BED data are presented at the establishment or the UI reporting unit level. Aggregating to the level of firm using EINs helps link firms that are operating in multiple states. Although these firms would have separate UI accounts for each state, they will have one EIN covering all of the establishments across the country. The EIN, however, is an imperfect way to

⁹ Some researchers also refer to this database as BLS Longitudinal Database or LBD (see Pivetz, Searson, and Spletzer, 2001).

¹⁰ See technical notes to the most recent BED press release (BLS, 2007c).

aggregate to the firm level. A multiunit firm can be associated with a cluster of one or more EINs. This distinction is not clear in the BLS data.

Other Variables. As with the QCEW, the BED also contains mailing and physical address and industry classification.

Major Limitations

A major limitation of this database is that it is not publicly available. Interested researchers need to apply for access to this database through the BLS research center.¹¹

Another limitation of this data set is that a significant change in the reporting does not allow for direct comparisons of the records before and after 1991. Prior to 1991, employers provided data on a reporting-unit basis, which was at the county level. While they often did, these units did not necessarily coincide with establishments and might have included several establishments within a county as long as they were conducting the same type of industrial activity.

While these data allow for identification of openings, closings, expansions, and contractions of establishments, there is likely to be a lag time in correctly identifying closings due to state agencies' tendency to impute employment figures for as many as two quarters after zero employment is reported.

What Can Be Done with the Data

Data from the BED program can help provide a dynamic picture of local labor markets. It is an important new database that includes the universe of longitudinal establishment data. As such, these data make it possible to study within-establishment changes over time and across location. For example, Faberman (2001, 2002, 2004) used these data to study different aspects of job creation and job destruction, job flows, and labor dynamics. In addition, these data can be used to answer basic questions about changes in establishment and firm sizes over time. For instance, it is possible to study growth in firms and establishments by employment size categories recorded at different points in the year. This information is currently unavailable and could change inferences based solely on March employment or average annual employment. These changes can be connected to various state-, county-, or even city-level policies, so that researchers can use geographic variation in policies over time to examine their effects.

Researchers can also use aggregated tables that provide statistics derived from the BED microdata. These series provide information about gross job gains and losses, expansions and contractions, and job openings and closings by establishment size categories. This information is publicly available, aggregated to the state level. These records are provided at both firm and establishment levels. For a detailed description of how these statistics were developed, see Butani, Clayton, et al. (2006).

¹¹ For the list of available data sets and application rules, see BLS (2007b).

CURRENT EMPLOYMENT SURVEY

Data Collection Method and Coverage

BLS conducts several surveys that provide monthly information on labor market conditions, including the Current Employment Statistics program (CES). As part of the CES, BLS cooperates with the State Employment Security Agencies (SESAs) to collect monthly, establishment-level data on employment, hours, and earnings. SESAs obtain this information from a sample of about 160,000 firms and government agencies monthly (about 400,000 establishments). All firms with 1,000 or more employees are asked to participate in the survey, as is a sample of firms across all smaller employment sizes. The CES survey draws most of its sample from the UI administrative records. The sample excludes all agriculture, private households, and self-employed workers. In addition, it includes some jobs not covered in QCEW. For example, it includes employees of railroads and religious organizations, as well as some other non–UI-covered jobs. As of 2003, the sample design is a stratified, simple random sample of establishments from the BLS’s Longitudinal Data Base, with strata defined by state, industry, and employment size.

Main Variables

Establishment and Firm Size Variables. The survey obtains information from each establishment on the total number of employees on their payroll and hours worked.

Other Variables. In addition to the total number of employees, the CES includes information on the number of female employees and the number of production workers. It also gathers information on several kinds of changes in employment or wages as well as reasons for these changes. For employment measures, data are collected on seasonal changes, short-term projects, layoffs, strikes, temporary shutdowns, or internal reorganization. For wage changes, data are collected on shifts in the wage rate, change in the quality of personnel, changes in the hourly pay or incentive pay, and overtime. In addition, data are collected on location and industry specialization.

Major Limitations

This survey has several limitations. First and foremost, the survey is not publicly available but is available only for use by approved researchers at BLS. From these data, a large number of employment, hours, and earnings series in considerable industry and geographic detail are prepared and published each month, although none is available by the categories of establishment or firm size.

Another issue is that the data set is not longitudinal, although the sample design uses rotating samples of establishments that follow some firms for about four years. The possibility of linking these establishments over time has not yet been discussed. In addition, survey nonresponse may be an issue. For example, Copeland (2003) notes that, depending on the industry, complete responses are obtained for only about 47–57 percent of the sample.

In addition, there was a major redesign to the CES survey in 2003, which might make comparisons over time problematic. In particular, the sample design was changed from an historical quota sample design to a probability sample. Details of the change are described in Butani, Stamas, and Brick (1997) and Morisi (2003).

What Can Be Done with the Data

To date, most of the work that used CES did not consider issues relating to firm size. Also, most of the research relies on the publicly available national level data not stratified by establishment size. These studies are often concerned with employment and wage trends in selected industries (see, for example, Krantz, 2002; Hatch, 2004; and Strople, 2006). Other studies are also concerned with the trends in hours of work (Kirkland, 2000; Hetrick, 2000; Kropf and Getz, 1999). None of these studies, however, considers these issues from the point of view of small businesses.

If researchers obtain access to the full CES, they can use the information to examine a variety of policy-relevant issues. Even though CES is based on a survey sample rather than on administrative records, it provides information that is not available in QCEW and that could be important for tracking economic development. In particular, the CES provides information on reasons for changes in numbers of employees and the number of women and production workers. Some researchers may use CES to examine effects of various policies on the development of the firms. For example, these data allow researchers to examine whether firms react to policies by increasing the work hours of existing employees or by hiring additional workers.

NATIONAL COMPENSATION SURVEY

Data Collection Method and Coverage

BLS conducts a number of surveys aimed at obtaining detailed information about compensation and benefits. Recently, a number of compensation surveys were combined into the National Compensation Survey (NCS) program.¹² Since 2000, the NCS survey has collected information that was previously collected using the Occupational Compensation Survey (OCS), the Employment Cost Index (ECI), and the Employee Benefit Survey (EBS). The NCS provides detailed measures of occupational earnings, compensation cost trends, benefits, and detailed plan provisions. The survey is fielded throughout the year. The reference month for average payroll information is July, while other information reflects the establishment's practices on the day the survey is completed.

The NCS is fielded every year and includes about 18,000 establishments, which are selected in a three-stage design. In the first stage, regions of interest are selected. In the second stage, establishments within these

¹² For a description and further information about the NCS program, see the program's Web site (BLS undated[b]), and the *BLS Handbook of Methods* (BLS, 1997).

areas are chosen, with the sample frame stratified by ownership and industry. The list of establishments is derived from the QCEW records. Before 1999, the NCS included only establishments with 50 or more workers. Since its beginning in 1999, the survey has covered establishments with one or more workers and state and local public agencies with 50 or more workers. Federal government, agriculture, and private households are excluded from coverage. According to the technical note to the NCS report, each sampled establishment was selected within a stratum with a probability proportional to its employment (BLS, 1997). In the third stage, a probability sample is taken of occupations within an establishment. Compensation for workers of specific occupations is collected (see BLS, 2006b).

Main Variables

Firm and Establishment Size Variables. The NCS data include variables on establishment employment and workers' variables including full-time and part-time status, hours worked, and wages.

Other Variables. Other worker-level variables are occupational details, bonuses, union status, and type-of-pay arrangements. Other establishment-level variables are full address, industry classification, and establishment characteristics. The NCS collects data on whether the firms provide different benefits and whether workers use those benefits. For example, the survey collects information on whether the firms provide life and disability insurance, retirement benefits, or defined contribution plans and indicates whether the workers elected to participate in those plans. It also collects information on whether the establishment provided various paid leave benefits including paid vacations and holidays, jury duty leave, and paid military leave benefits. The NCS collects information on a variety of health benefits, including medical, prescription drug, dental, and vision care. It describes the health care plans and systems provided by the establishment, limitations on coverage, and whether employees are using those plans. In addition, the NCS program collects limited data on emerging types of benefits that establishments may provide. Those include educational assistance, employee wellness, adoption assistance programs, travel accident and long-term care insurance, supplemental unemployment, flexible workplace programs, fitness center benefits, on-site child care, and commuting subsidies.¹³

Major Limitations

There are several major limitations to using this survey for the analysis of small business affairs. The first limitation is that the data are not publicly available. Researchers need to apply for access to this database through the BLS Research Center.¹⁴ Some aggregate tables for localities, broad regions, and the nation by establishment employment size category are available (see BLS, undated[b]).

It may be problematic to compare records before and after 1999 due to a redesign of the survey. In particular, BLS integrated three different surveys into a single compensation survey. As a result, there were

¹³ For more information, see BLS (2006c).

¹⁴ For the list of available data sets and application rules, see BLS (2007b).

changes in how the data were collected from establishments of different size. Prior to 1999, establishments of different size were targeted in different years—data for medium and large establishments (more than 100 employees) were collected in odd years and for small establishments in even years. In addition, the sample is selected in such a way that larger firms are more likely to be surveyed.

What Can Be Done with the Data

NCS data have been used to examine factors that determine low-wage labor (Bernstein and Gittleman, 2003), incidents of provision of health benefits (Barsky, 2004), and trends in employer-provided prescription drug coverage (Dietz, 2004). Each of these topics could be addressed by establishment employment size categories. With these data, it also would be possible to evaluate the relationship between employer-provided health benefits and establishment or worker characteristics by establishment size.

CURRENT POPULATION SURVEY

Data Collection Method and Coverage

In addition to establishment-level data, one of the BLS' most widely used household surveys, the Current Population Survey (CPS), also can be used to identify some members of the small business universe. The U.S. Census Bureau conducts the CPS for BLS. The microdata from this survey are publicly available with coverage beginning in 1962. BLS uses the data to provide monthly estimates of the number of unemployed people in the United States. The CPS also serves as a vehicle for supplemental studies on subjects other than employment. In contrast to other surveys we have discussed, the CPS uses the household, rather than the establishment or firm, as its sampling unit. The survey nonetheless provides a comprehensive current source of information on the occupation of workers and the industries in which they work. However, the survey provides only a limited amount of firm-level information. For these reasons, the CPS could provide a particularly useful source of information about the self-employed and new entrepreneurs.

The CPS collects information on the labor force status of the civilian, noninstitutional population 15 years of age and older. The data are obtained from a sample of about 60,000 households. The CPS stands out among government surveys as having consistently very high response rates. The households are selected using a multistage, stratified, statistical sampling scheme.¹⁵ The data for all members of the household are recorded in separate records. The sample is selected to assess overall employment, unemployment, and the number of people in and out of the labor force. This sample includes categories of workers that are entirely or partly excluded from the QCEW program. For instance, people are classified as employed if they did any work at all as paid employees during the reference week; worked in their own firm, profession, or on their own farm; or worked without pay at least 15 hours in a family firm or farm. People are also counted as employed if they

¹⁵ Details about CPS methodology are provided in BLS (2002).

were temporarily absent from their jobs because of illness, bad weather, vacation, labor-management disputes, or personal reasons.

Researchers can take advantage of the panel structure of the CPS. Every housing unit in the CPS is interviewed for four consecutive months and then dropped out of the sample for the next eight months and brought back in the following four months. In most years, the observations can be linked over time using household identification number. However, these links can generate false positives because of nonresponse, migration, mortality, or recording errors. Algorithms are available to improve matching using changes in gender, race, age, and educational attainment. Trade-offs of these algorithms are discussed in Madrian and Lefgren (1999).

Main Variables

Firm and Establishment Size Variables. In addition to the demographic and labor force information mentioned below, the March demographic supplement of the CPS collects information on firm employment. The specific wording of the question in the March supplement is, “NOEMP—Counting all locations where this employer operates, what is the total number of persons who work for [the respondent’s] employer?” Response categories are under 10, 10–24, 25–99, 100–499, 500–999, and 1,000+.

Other Variables. For each member of the household, the basic CPS monthly survey collects demographic information, such as age, sex, race, marital status, veteran status, Hispanic origin, immigrant status, educational attainment, and family structure. In addition, information is collected on labor market outcomes: employed, unemployed, and whether searching or not searching for work. The reference period for labor market information is defined as the seven-day period (from Sunday through Saturday) that includes the 12th of the month. Respondents are also asked questions about class of worker (private, government, self-employed, without pay, and never worked), hours worked in the reference week, occupation, industry type, reasons for working part time, and reasons for lack of employment. Respondents in the outgoing rotation panel (those households that are in the panel for month number 4 or number 8) are asked about earnings in their main job. The self-employed are usually asked a set of questions similar to those asked of the other workers. New business owners can be identified by matching two of the multiple time points at which an individual is included in the survey, either across months or across years and by flagging those who are business owners at the second but not the first time point. In addition, the March demographic supplement survey collects information on health insurance and other noncash benefits provided.

Other supplements on selected topics are included for most months. These supplemental topics are often repeated in the same month from year to year. For example, biennial September supplements gather detailed information for veterans, their service-connected income, effects of a service-connected disability on current labor force participation, and participation in veterans’ programs. Biennial February supplements collect information about type of employment arrangement workers have on their current jobs and other characteristics of the current jobs such as earnings, benefits, longevity, employee satisfaction rates, and

expectations. Additional supplements collect information about computer and Internet use at home and at work, job tenure and occupation mobility, adult education, and health and pension coverage.

Major Limitations

The CPS has several limitations, the most important of which for our purposes is that it contains limited information about firms. The CPS is a household survey designed to collect unemployment data, and very little information is collected about firms where respondents work or that they own. Other possible disadvantages of the CPS are that it covers only a two-year longitudinal panel and loses a significant portion of the sample due to matching algorithms or moving of households. In addition, a major redesign of the survey instrument in January 1994 might constrain comparison of some series before and after 1994.

What Can Be Done with the Data

The overall advantages of the CPS are large sample sizes, long time series, quick access to timely data, a very large built-in comparison group of nonentrepreneurs, and a wide range of topics in the supplements. The coverage of the CPS makes it an important tool for the analysis of the self-employed and new entrepreneurs. A number of papers use the CPS for the analysis of issues that are relevant for small businesses. For example, Berger (1999) used March CPS data to examine the distribution of low-wage workers by firm employment size categories and investigates effects of the minimum wage. Labor economists have used the CPS extensively for the analysis of the relations between employment size of a firm and wages (see Idson and Feaster, 1990; Mellow, 1982; Bowlus, Kiefer, and Neumann, 1995; Pearce, 1990; Card, 1996; Brown and Medoff, 1989; Antos, 1983; Hirsch and Schumacher, 1998; Weiss and Landau, 1984; Evans and Leighton, 1989). Other studies include analyses of the prevalence of formal on-the-job training (Loewenstein and Spletzer, 1997), factors that explain differences in turnover between large and small firms (Even and Macpherson, 1996), effects of health insurance on hours worked (Cutler and Madrian, 1998; Gruber and Poterba, 1994), effects of employment protection (Oyer and Schaefer, 2002), earnings by racial or ethnic characteristics (Agesa, Agesa, and Hoover, 2001; Carrington, McCue, and Pierce, 2000; Trejo, 1997), patterns of entrepreneurship (Evans and Leighton, 1989), gender differences in earnings (Sorensen, 1990; Macpherson and Hirsch, 1995), evidence of labor market cycles for the self-employed (Carrington, McCue, and Pierce, 1996), transition between full-time work, part-time work, and retirement (Peracchi and Welch, 1994), worker compensation (Hirsch, Macpherson, and DuMond, 1997), patterns of self-employment among older U.S. workers (Karoly and Zissimopoulos, 2004), and access to computers and the decision to become self-employed (Fairlie, 2005).

The CPS can be used to support further studies of self-employment and entrepreneurship, including patterns of health insurance coverage, human capital, and education among the self-employed. The CPS can also be used to examine patterns of self-employment and entrepreneurship among recent immigrants or other demographic groups of interest.

Chapter Three

CENSUS BUREAU DATA SOURCES

The U.S. Census Bureau is the leading source of quality data about the nation's people and economy. In addition to the widely known decennial population census, the Census Bureau collects and maintains a number of data sets that provide information on businesses. These include the Business Register (BR), Business Information Tracking System (BITS), Integrated Longitudinal Business Database (ILBD), Economic Census (EC), Company Organization Survey (COS), Survey of Minority-Owned Business Enterprises (SMOBE), Survey of Women-Owned Business Enterprises (SWOBE), and Survey of Business Owners (SBO). These data sources vary in terms of how comprehensive they are, how often they are updated, and whether it is possible to create longitudinal series on the basis of these data.

The BR provides a comprehensive roster of each known establishment and company. It covers both employer and nonemployer businesses. The data set draws its information from administrative records (e.g., IRS and Social Security Administration [SSA] records) and surveys (e.g., the EC and the COS). The detailed cross-sectional data are available for each year since 1975. The BR contains information on establishment and firm employment, payroll, revenues, full address, firm affiliation, and industry classification.

County Business Patterns (CBP) data aggregate BR data to the level of county for establishments with paid employees. Some industries that are represented in the BR data are excluded from the CBP roll-up. In particular, the coverage excludes some agricultural industries, railroads, postal service, private households, large pension, health, and welfare funds and public administration. Yearly tables are provided for data from 1964 to the present. The CBP tables are often used to derive denominators for employment as well as the number of establishments in a particular establishment-size category by county.

Two longitudinal databases based on the BR data are available. These longitudinal data sets are particularly useful to researchers who want to study the emergence and dynamics of small business because they allow for an examination of entry, exit, and gross job flows by establishment or firm employment size. The LBD is created by linking information on all establishments included in the BR across years. The BITS links information longitudinally for those establishments included in the annual CBP data. There are several key differences between the LBD and the BITS. First, the BITS has more restrictive coverage of industries, since it uses the same coverage as the CBP. In addition, the LBD has longer panels. While the BITS goes back only until 1988, the LBD links establishments back to 1976. On the other hand, the BITS currently includes both firm and establishment data, while both levels of data are not yet available for the LBD.

Two census survey efforts feed into the underlying BR data: the COS and the EC. These surveys are used to gather up-to-date company affiliation, location, and operating information for establishments that are part of multiestablishment companies in the BR. Such information may be relevant for researchers interested in special topics. The COS surveys all multiple-unit firms with more than 250 employees every year and

smaller multiple-unit firms on a rotating basis. The EC covers all establishments of multiunit companies, all single-unit employers larger than the industry size cut-off (usually three employees) and a sample of smaller single-unit employers. It gathers data on revenue, payroll, address, and ownership type, as well as other sector- and industry-specific information.

Three related surveys, the samples of which are drawn from the BR data frame, are SMOBE, SWOBE, and SBO. These surveys supplement the BR and related databases with more detailed information on business owners. These data can be linked with firm size information from the BR to compare characteristics of owners of small versus large businesses and to understand how key features of business operations for large and small businesses.

Current efforts in the Bureau of Census are directed at developing integrated databases that include employer and employee characteristics by extending LBD along two dimensions. These data have great potential for studying dynamic changes in establishments and firms and connecting these to owner and worker characteristics. The first extension integrates nonemployer data, making it possible to track transitions to and from employer to nonemployer status.¹⁶ The data set including both employee and nonemployee businesses is the ILBD. The second extension is to include information from the Longitudinal Employer Household Dynamics (LEHD) files, which provide person and business identifiers for all workers and businesses covered by unemployment insurance in 30 states. The person identifiers can then be used to match workers to information in other person-level census products, such as the Survey of Income and Program Participation (SIPP) or the long-form census.

A major limitation of all these data sources is that they are not publicly available. In most cases, researchers can apply for access to the microdata at the Census Research Data Centers (RDCs).¹⁷ The Census Bureau publishes tables based on the underlying data, and, in many cases, researchers can request that the Census Bureau produce additional tables summarizing information on specific variables of interest. Researchers can combine many of these sources to provide more detailed information about the small-business universe.

STANDARD STATISTICAL ESTABLISHMENT LISTING OR BUSINESS REGISTER

Data Collection Method and Coverage

The Standard Statistical Establishment Listing (SSEL), also known as the Business Register (BR), contains records for each known establishment and company that is located in the United States. The detailed cross-sectional data are available for each year since 1975, although the frequency of updating the data varies depending on the industry. The BR systematically incorporates information about existing and new firms and

¹⁶ In 2002, there were approximately 16 million nonemployer businesses. Approximately 14 million do not have EINs but are uniquely identified by the owner's social security number (SSN); the other 2 million have EINs (Davis et al., 2006).

¹⁷ For a detailed description, see U.S. Bureau of the Census (undated[a]). A criterion for use of the data is that the proposed research must benefit the Census Bureau.

establishments from a number of administrative sources, censuses, and surveys.¹⁸ The first source is payroll tax information provided by the IRS, which provides employment and payroll information for existing firms. The second source is for new firms and is provided by the SSA. The data come from applications for EINs completed by all new employers. The Business Establishment List, separately created and maintained by BLS from unemployment insurance administrative records, is used to fill in industry classifications (SIC codes) for establishments otherwise missing this information. The BR information is regularly updated (so accurate information is available in each cross section) using the COS, an annual survey of manufacturers, and the EC.

As described on the Census Bureau Web site (U.S. Bureau of the Census, 2006a), this data set covers establishments of all domestic employer and nonemployer businesses (except private households and governments) and establishments that are parts of multiestablishment firms. The cross section includes 180,000 multiunit companies, representing 1.5 million affiliated establishments, 5 million single-establishment companies, and nearly 14 million nonemployer businesses. Although both BR and QCEW serve primarily as lists of existing establishments for the Bureau of Census and BLS, respectively, these listings have important differences in collection, scope, data definition, and reference periods.¹⁹ For example, there are important differences between the BR and the QCEW in terms of coverage. As described in Becker et al. (2005), when compared to BR, the QCEW does not include nonemployer establishments in its listings. In addition, the QCEW does not include records for establishments that are not subject to state and federal unemployment insurance programs. On the other hand, the QCEW covers such sectors as agricultural production, pensions and other funds, trusts, postal workers, and private households, while the BR does not. The BR also excludes from coverage most government establishments.

This data source is used to create the Statistics of U.S. Businesses (SUSB), the CBP, and ZIP Code Business Patterns (ZBP), all of which are useful sources because the aggregate statistics they contain are made publicly available at fine geographic and industry levels and by establishment employment size.

Main Variables

Firm and Establishment Size Variables. For each establishment record, the BR has information on the number of employees, payroll, and receipts. The employment and payroll data are continuously updated using various administrative sources. The data also include identifiers that can be used to connect establishments to firms and units over time. Those are Census File Number (CFN), EIN, and Permanent Plant Number (PPN).

Other Variables. Also available for each establishment record are full address, company affiliation information, and industry classification.

¹⁸ A description of the BR is available at U.S. Bureau of the Census (2006a).

¹⁹ For more details about each of these dimensions, see Becker et al. (2005).

Major Limitations

One of the major limitations of this data set is that it is not publicly available. Researchers can apply for access to the microdata at the census RDCs.²⁰ In addition, aggregated tables from the data are available through CBP. Additional tables also can be requested and are provided at a cost commensurate with the work involved in creating the table.

What Can Be Done with the Data

The BR provides underlying records that form a number of statistical databases within the Census Bureau. For example, the LBD and BITS connect some of the records from the BR over time to create longitudinal databases of establishments. In addition, the BR serves as a sampling frame for a number of surveys conducted by the federal government.

LONGITUDINAL BUSINESS DATABASE

Data Collection Method and Coverage

The Center for Economic Studies of the Census Bureau links the annual snapshot files from the BR over time to create the LBD. Jarmin and Miranda (2002) describe in detail how the LBD is constructed. In particular, the matching algorithm relies on available numerical identifiers as well as information about name and address of the firm to make linkages over time. Three numerical identifiers are available in the LBD. The Census Bureau uses the CFN to identify establishments in ECs and surveys. However, this number can change over time due to changes in ownership status, in single- and multiunit status, or in the legal form of an organization. The Census Bureau tracks changes in the CFN. Another identifier is the PPN. This number was introduced to BR in 1982 and is designed to stay unchanged as long as the establishment remains active at the same location. The other available identifier is the EIN. This is a taxpayer identifier assigned by the IRS. As described in Jarmin and Miranda (2002), longitudinal matching first uses information in PPN, but, if there are no matches, the matching is done using CFN, and, after that, by EIN and possibly tracking changes to CFN. Matching is also augmented using the name and address of the firm.

The resulting LBD covers nearly all of the nonfarm private economy, as well as some public sector activities from 1975 to the present. It includes about 4.5 to 7.1 million records per year, for a total of almost 24 million unique establishments from 1975 to the present. It excludes establishments with zero annual payrolls.

Main Variables

Firm and Establishment Size Variables. The LBD draws its main variables from the BR. Those include establishment and firm employment, payroll, and revenues.

²⁰ For a detailed description, see U.S. Bureau of the Census (undated[b]).

Other Variables. Similar to the BR, the LBD also includes full address, firm affiliation, and industry classification. In addition, the LBD includes information on the enterprise age and tenure. There are also plans to include variables on ownership status and ownership changes.

Major Limitations

Like other microdata from the Census Bureau, this data set is not publicly available. Researchers can apply for access to the microdata at census RDCs.²¹ In addition, the accuracy of the firm-to-establishment links varies over time; quality of the links declines after an EC and then improves again with the next census (Jarmin and Miranda, 2002).

What Can Be Done with the Data

The LBD is useful to researchers who want to examine entry, exit, and gross job flows by establishment or firm employment size. The data allow researchers to study changes over a long time frame within establishments. In the past, the data set was used to examine entry and exit of firms in specific industries (Jarmin, Klimek, and Miranda, 2004) and establishment and employment dynamics (Foster, 2003). This data set can be connected to other Census Bureau products.

COUNTY BUSINESS PATTERNS

Data Collection Method and Coverage

CBP is one of the programs that provide geographic aggregates of microlevel establishment data. The CBP aggregates data from the BR to county level according to industry and establishment employment size category (see U.S. Bureau of the Census, 2006a). However, it does not provide data by firm size categories. Yearly tables are provided for data from 1964 to the present.

The CBP covers all establishments with paid employees included in the BR, although it excludes some industries. In particular, the CBP excludes crop and animal production; rail transportation; National Postal Service; pension, health, welfare, and vacation funds; trusts, estates, and agency accounts; private households; and public administration. The CBP also excludes most government establishments.

Main Variables

Firm and Establishment Size Variables. The CBP provides aggregate tables for employment during the week of March 12, total number of establishments, first quarter and annual payroll by industry, and establishment employment size class. The CBP uses nine different employment size categories: one to four, five to nine, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, and 1,000 or more employees. The data on total employment and payroll are suppressed whenever they would disclose the operations of an

²¹ For a detailed description, see U.S. Bureau of the Census (undated[b]).

individual employer. In addition, the ZBP data provide the number of establishments by industry codes and establishment employment categories for each ZIP code.

Other Variables In addition to the variables above that may be used to determine establishment size, the CBP and ZBP also classify information by location, specifically county or ZIP code, and industry.

Major Limitations

Although this product is publicly available, it provides only aggregated tables. These might be less useful in many instances in which firm- or establishment-level data are necessary. In addition, the CBP does not provide information by firm employment size, only by the establishment size. Therefore, the researchers should be careful when making inferences about firm behavior in the industries in which many firms consist of multiple establishments.

Another major limitation of this product for study of the small businesses is that CBP does not cover establishments without paid employees. These so-called nonemployer establishments include small family-operated businesses and small businesses without paid employees. These establishments, however, encompass significant amount of entrepreneurial activity. As a result, the aggregated tables may not provide a reliable picture of the entrepreneurial activity in the industries that includes a sizable number of establishments without paid employees. The researchers may avoid this problem when they combine records from CBP with the Census Bureau nonemployer statistics.²²

What Can Be Done with the Data

The CBP data set is a standard reference source of local economic data. The CBP tables are often used to derive denominators for employment and number of establishments in a particular establishment size category (numerator data typically come from other sources). These tabulations can be used to examine effects of state- and county-level policies on establishments of varying sizes and in different industries.

BUSINESS INFORMATION TRACKING SYSTEM

Data Collection Method and Coverage

Over the years, there have been several efforts to link records in the BR over time to create a longitudinal database of establishments.²³ The BITS, also known as the Longitudinal Enterprise and Establishment Microdata (LEEM), is one of the most recent efforts to create a longitudinal database. The Census Bureau has collaborated with the SBA's Office of Advocacy to create the BITS, which links establishment data from BR over time from 1989 to 2001 (and ongoing with a two-year lag) and within firms. As described in Acs and Armington (2005), the primary links over time are constructed using the CFN,

²² For information about census nonemployer statistics, see U.S. Bureau of the Census (2006b).

²³ Also see description of the LBD.

the Census Bureau identification number for establishments in the EC. However, this number would not connect establishments that changed ownership or organizational structure between years. These establishments are matched using a Permanent Plant Number (PPN), an establishment identifier that does not change over time; EIN; and other establishment attributes like name, address, zip code, and industry codes. Robb (1999) and Acs and Armington (1998) provide documentation for the BITS data.

There are several key differences between the LBD and the BITS. First, the BITS has more restrictive coverage of industries. It uses the same coverage as the CBP. In particular, the coverage excludes some agricultural industries, railroads, postal service, private households, large pension, health and welfare funds, and public administration. In addition, the LBD has longer panels. While the BITS goes back only until 1988, the LBD links establishments back to 1976. On the other hand, BITS currently includes both firm and establishment data, while both levels of data are not yet available for the LBD.

Main Variables

Firm and Establishment Size Variables. The BITS includes establishment-level information on employment (missing for 15 to 18 percent) and annual payroll. The records also include firm-level information on aggregate employment and payroll.

Other Variables. Other establishment-level data are location information including state, MSA, city and place codes, start year, and industry code. Other firm-level information includes primary industry and primary location (based on the establishment with the largest number of employees).

Major Limitations

The BITS data are not publicly released, but researchers can apply to use the data at Census Research Data Centers (RDCs). Researchers are encouraged to work with staff at one of the RDCs to develop a research proposal, which is likely to take a minimum of six months to be reviewed and, at times, substantially longer. The proposed research must further the interests and goals of the Census Bureau.²⁴

The largest changes in the BITS data occur in and just before EC years, when industry classifications are clarified and changes in establishments within firms are recorded.

What Can Be Done with the Data

Using this longitudinal database, it is possible to identify establishment births, deaths, expansions, and contractions. Most census products can be connected to each other using an EIN or PPN. Therefore, information from the other census products could be used to identify a type or category of establishments, and BITS can be used to track those establishments over time. There are several studies that used BITS data to examine issues that may be important for small businesses. The data were used to examine the persistence of new jobs (Acs and Armington, 2004); job flow dynamics (Acs and Armington, 1999; Acs, Armington, and

²⁴ For more information, see Acs and Armington (2005).

Robb, 1999); survival of firms in various industries, including start-ups (Headd, 2001; Boden, 2001); and mergers and acquisitions (White, 2002; U.S. Small Business Administration, 1998).

ECONOMIC CENSUS AND COMPANY ORGANIZATION SURVEY

Data Collection Method and Coverage

Two census products central to the BR, the BITS and the LBD, are the Company Organization Survey (COS), also known as the Report of Organization Survey, which is fielded annually except in Economic Census years, and the Economic Censuses (ECs), which are fielded in years ending in two and seven. This information is used to maintain up-to-date company affiliation, location, and operating information for establishments that are part of multiestablishment companies in the BR. The COS surveys all multiple-unit firms with more than 250 employees every year and smaller multiple-unit firms on a rotating basis. It maintains information on the organizational design and employment of multiunit firms. In particular, companies identify changes in their establishments due to sale or closure or the start-up or acquisition of new establishments. The companies are asked to indicate controlling interests held by other domestic or foreign-owned organizations. Law mandates completion of both the EC and the COS.

The EC covers all establishments of multiunit companies, all single-unit employers larger than each industry size cutoff (for most industries about three employees), and a sample of small employers with fewer employees than the industry employment size cutoff in most industries except agriculture and government (see U.S. Bureau of the Census, 2006a). Its purpose is to provide comprehensive statistics about establishments and their activities. It covers all domestic nonfarm business establishments other than those operated by the government.

Main Variables

Firm and Establishment Size Variables. The EC collects data on establishment employment in the pay period that includes March 12, total revenue, and annual and first quarter payroll.

Other Variables. The EC also includes establishment full address, organizational form, and type of ownership. Additional information is collected for some sectors and industries. Those include Census of Finance, Insurance and Real Estate (CFI), Census of Manufactures (CMF), Census of Retail Trade (CRT), Census of Transportation, Communications, and Utilities (CUT), Census of Wholesale Trade (CWH), and Census of Services (CSR).

Major Limitations

A major limitation of the EC and COS is that these data sets are not publicly available. Researchers can apply to use the data at census RDCs.²⁵ The Census Bureau releases aggregated tables from the EC for

²⁵ For more information, see U.S. Bureau of the Census (undated[a], undated[b]).

industry data by firm employment size. The Census Bureau releases most tables from an EC three to four years after it was conducted.

What Can Be Done with the Data

As we have stated, the EC and the COS serve as two of the main surveys that add data to the BR. The type of ownership and organizational form data could be compared across firms of different size using employment-, payroll-, and revenue-based size definitions. In addition, researchers can make use of the detailed industry information collected in the ECs. For example, Garicano and Hubbard (2005a, 2005b) used data from the 1992 CSR to study specialization within and among law firms.

SURVEY OF WOMEN- AND MINORITY-OWNED BUSINESS ENTERPRISES, CHARACTERISTICS OF BUSINESS OWNERS SURVEY, AND SURVEY OF BUSINESS OWNERS

Data Collection Method and Coverage

Four related surveys, the samples of which are drawn from the BR data frame, are the Surveys of Minority- and Women-Owned Business Enterprises (SMOBE, SWOBE), Characteristics of Business Owners (CBO) survey,²⁶ and the Survey of Business Owners (SBO). These surveys are conducted once every five years in conjunction with the EC. The SMOBE and SWOBE were carried out in 1992 and 1997 and the CBO and SBO in 1992 and 2002. The U.S. Code Titles 13 and 26 authorize these data collections and provide for mandatory responses (see U.S. Bureau of the Census, 2006a).

These surveys supplement the BR and all related products with more detailed information on business owners. The 2002 SBO samples from all businesses that file tax forms as individual proprietorships, partnerships, or corporations with receipts of \$1,000 or more. Major industries that are excluded from the survey include agricultural production, domestically scheduled airlines, railroads, the U.S. Postal Service, private households, and some nonprofit organizations. The sampling methodology includes a notable method of sorting business into gender, racial, and ethnic categories based on a series of probabilities obtained from additional sources.

Main Variables

Firm and Establishment Size Variables. The surveys do not collect any establishment or firm size data; however, the data can be linked to the size information from the BR or EC using identifiers available in the confidential data.

Other Variables. These surveys collect detailed information on business owners, including age, education level, sources of financing, gender, race, and ethnic background. The 2002 SBO also collects

²⁶ Headd (1999) provides documentation for the 1992 CBO database.

information on veteran status, service, disability status, types of customers (federal government, local government, consumers) and workers, home-based firms, family-owned firms, and sources of financing for capital improvements or start-up (see U.S. Bureau of the Census, 2002b).

Major Limitations

One major limitation of these surveys is that they are not publicly available. Researchers need to apply for access to these surveys to the census RDCs.²⁷ These surveys are available as part of the EC (see U.S. Bureau of the Census, undated[a]). The Census Bureau releases detailed tables from the SBO approximately four years after the data are collected. These aggregated tables from the survey include information on number of firms, sales and receipts, paid employees, and annual payroll. These tables are presented by geographic area, industry, firm size, and legal form of the organization (U.S. Bureau of the Census, 2002c).

In addition, between 1992 and 1997, there were changes in the survey methodology that might limit direct comparisons between years. In particular, changes were made to the target population (the universe was extended to include different types of corporations) and to the definition of business (in 1997, all operations under the same ownership were defined as one company even though they might have had different EINs) (U.S. Bureau of the Census, 2002a).

What Can Be Done with the Data

These data sources provide a rich and unique set of information on the characteristics of small and large business owners and their sources of financing. The researchers often use confidential records from the CBO to study factors that affect entrepreneurship. For example, Fairlie and Robb (2005) use the 1992 CBO survey to examine how family and human capital factors affect sales, profits, employment size, and survival probabilities of businesses owned by minorities. Holmes and Schmitz (1992) used the 1982 CBO survey to examine how failure and sale of the small business depends on the characteristics of the managers. Bates (1995) used the 1987 CBO survey to examine the effect of state and local government managerial, technical, and procurement assistance on survival of small businesses. The National Research Council (2005) issued a report using the SWOBE along with other data sources focusing on federal contracting and women-owned small businesses.

One of the benefits of these surveys is that they can be connected to the other census products that have longitudinal information on employment and revenues. Using the 2002 SBO, one can also examine characteristics of firms that have federal and local public agencies as their main consumers.

²⁷ The approval for using a survey should be obtained from the Center for Economic Studies at the Census Bureau and also from the IRS, which provides records for survey sample frame.

INTEGRATED LONGITUDINAL BUSINESS DATABASE AND LONGITUDINAL EMPLOYER-HOUSEHOLD DYNAMIC PROGRAM

Data Collection Method and Coverage

Current efforts in the Bureau of Census are directed at developing integrated databases that include employer and employee characteristics, as reported by John Haltiwanger²⁸ at the Kauffman Symposium on Entrepreneurship Data held November 10–11, 2004. The focus of these efforts is on extending the LBD along two dimensions. The first extension integrates nonemployer data making it possible to track transitions to and from employer and nonemployer status.²⁹ The second extension is to include information from the Longitudinal Employer Household Dynamics (LEHD) files, which provide person and business identifiers for all workers and businesses covered by UI in 30 states. The person identifiers can then be used to match workers to information in other person-level census products, such as SIPP or the long-form census.

The data set including both employee and nonemployee businesses is called the Integrated Longitudinal Business Database (ILBD). The ILBD combines administrative records and survey-based data for virtually all employer and nonemployer business units in the United States. The data integrate the LBD, discussed above, with the nonemployer data (Davis et al., 2006). In 2000, there were roughly 14 million nonemployer businesses. The census defines a nonemployer business as one that has no paid employees, has annual business receipts of \$1,000 or more (\$1 or more in the construction industries), and is subject to federal income taxes.³⁰ These are usually self-employed individuals operating a very small, unincorporated business. For small nonemployer businesses, the Census Bureau obtains administrative records from individual income tax returns. These small nonemployer businesses are likely not to be assigned an EIN and are tracked using the self-employed individual's SSN. This new integrated data permits analysis of the movement of people between employer and nonemployer universes.

The LEHD connects establishment data from LBD to household data. The employer portion of this integrated database includes records for about 4 million establishments from more than 20 states. In his presentation at the 2004 Kauffman Symposium on Entrepreneurship Data, Haltiwanger encouraged researchers to persevere in making use of these rich data sources. For more information, see Davis et al. (2006).

Main Variables

Firm and Establishment Size Variables. The combined data set includes establishment-level information on monthly employment and quarterly wages.

²⁸ John Haltiwanger is a professor of economics at the University of Maryland and research associate at the Center for Economic Studies at the Census Bureau. He also was chief economist at the Census Bureau from 1997 to 1999.

²⁹ In 2002, there were approximately 16 million nonemployer businesses. Approximately 14 million do not have an EIN but are uniquely identified by the owner SSN, and the other 2 million have an EIN (Davis et al., 2006).

³⁰ See definition at U.S. Bureau of the Census (2005).

Other Variables. The data also include detailed location, industry, and workforce composition and worker turnover from 1990 to 2004 (years covered vary by state) at the establishment level. These records are extracted from the UI data and can be connected to other Census Bureau databases. In addition, many employer-level measures are created from longitudinally integrated person- and establishment-level data. The data include worker and job flows, including new hires, separations, job creation, job elimination by age and gender of workforce; worker composition by gender and age; worker compensation for stocks and flows by gender and age; and dynamic worker compensation summary statistics for stocks and flows by gender and age. This file also contains identifiers used by state ES-202 and UI systems as well as by federal EINs. A description of the data is presented in LEHD (2002).

Major Limitations

A major limitation of this data set is that it is not publicly available. Researchers can apply for access to microdata at the Census Research Data Centers.³¹

What Can Be Done with the Data

This data set has great potential for studying dynamic changes in establishments and firms and connecting these to owner and worker characteristics. The data set includes more data elements than is found in business owner surveys and covers an important part of the small business universe, such as nonemployer businesses. Current research is using the LEHD to study the impacts of new technologies on firms and workers (Abowd, Haltiwanger, Lane, and Sandusky, 2001); to measure the relationships between human capital and a firm's technology (Abowd, Haltiwanger, Jarmin, et al., 2002); and to examine the relationship between employer-provided health insurance, worker mobility, and wages (Stinson, 2002). The data are also used in the recent book *Economic Turbulence* (Brown, Haltiwanger, and Lane, 2006) that examines the impact of economic volatility on workers and businesses. For other papers that used LEHD, see U.S. Department of Labor (2004).

³¹ For a detailed description, see U.S. Bureau of the Census (undated[b]).

Chapter Four

OTHER GOVERNMENT SOURCES OF DATA

The census and BLS data described in the preceding chapters have distinct strengths and weaknesses. Because the data are derived from administrative records or compulsory surveys, they are extremely comprehensive and gathered in a relatively consistent way over time. As a result, they are useful for providing a cross-sectional overview of business characteristics, making comparisons among businesses of different sizes and also for examining changes in such characteristics over time. Researchers have some ability to link records over time and use these data to study firm dynamics, entry, and exit. However, the fact that most of these data are derived from administrative records also means that the type of information available is limited to fairly basic characteristics. This chapter describes several government efforts to survey firms to gather more detailed information on specific topics.

The survey efforts described in this chapter address business finance, health insurance, and medical expenditures. They provide detailed information from a random sample of businesses, including small businesses. As such, the data can support research on specific topics that compares large and small businesses. In general, the data are available to researchers, although researchers may be required to submit a proposal and pay a fee to use the data. Survey response rates can be quite low, and data are not always collected on a regular or continuing basis. This may compromise the generalizability of any research results and limit the ability to engage in longitudinal analysis.

SURVEY OF SMALL BUSINESS FINANCES

Data Collection Method and Coverage

In addition to these census and BLS sources of data, the Federal Reserve Board sponsors a survey of small firms. The Survey of Small Business Finances (SSBF) was conducted in 1987, 1993, 1998, and 2003. The SSBF contains information from more than 3,500 firms with fewer than 500 employees. It oversampled African-American-, Asian-American-, and Hispanic-American-owned firms. The sampling frame for the SSBF is the Dun's Market Identifier (DMI) file, which is described below. Unlike most of the government data sources described here, full public data sets for each of the first three years of SSBF are available from the Federal Reserve. This survey appears to have been well designed and the sampling plan and implementation are well documented.

Main Variables

Firm and Establishment Size Variables. This SSBF includes information on firm employment and income and balance sheets of the firm.

Other Variables. The survey also includes a number of owner characteristics and detailed information on the use of financial services. For instance, it collects information on the experience of small businesses with credit applications and credit access. The balance sheet section of the questionnaire provides a detailed description of the assets and liabilities of the firm, as well as information on the firm's credit history. This includes characteristics of profits, sales, labor productivity, and other measures of firm success. The questionnaire is extensive and includes information on a range of variables, from the ways in which computers are used in the business to what business owners consider the most important problem facing their business.

Major Limitations

Users of the survey should keep in mind that the survey was voluntary, and the response rate was approximately 33 percent. The survey also provides no information at the establishment level.

What Can Be Done with the Data

Researchers have used this survey to examine financial constraints that firms face (Lel and Udell, 2002; Robb, 2002); adoption of computers (Bitler, 2001); use of financial services (Bitler, Robb, and Wolken, 2001); borrowing experience by gender, race, and ethnicity of firm owners (Coleman, 2002a, 2002b, 2003); and the decision to become a public firm (Helwege and Packer, 2003). A description of the studies that used SSBF is provided at the Federal Reserve Web site (see Federal Reserve Board, 2006).

NATIONAL EMPLOYER HEALTH INSURANCE SURVEY

Data Collection Method and Coverage

A number of federal agencies collect data related to the provision of health insurance. For instance, in 1994 the Centers for Disease Control and Prevention conducted the National Employer Health Insurance Survey (NEHIS). The survey was designed to produce estimates on employer-sponsored health insurance data in the United States for establishments of different sizes.³² It served as a precursor to the Insurance Component of the Medical Expenditures Panel Study described below.

The NEHIS is a 1994 national probability sample survey of establishments, public sector entities, and self-employed individuals.³³ The sample was drawn from three sample frames—the Dun's Market Identifiers file for the sample of private establishments, the Census of Governments file for public sector entities, and the National Health Interview Survey to select self-employed individuals with no employees (NCHS, 1997). Among private sector establishments, 34,604 interviews were completed for a response rate of 71 percent.

³² See description at NCHS (2007).

³³ See description of the methodology at NCHS (2007).

Main Variables

Firm and Establishment Size Variables. The NEHIS collected information on establishment-level, firm-level, and state-specific total employment. Also included are establishment-level annual payroll.

Other Variables. The NEHIS contains detailed data on various aspects of employer-sponsored health insurance, types of plans offered to employees, and detailed characteristics of the offered plans (monthly premiums, employer and employee contributions to premiums, deductibles, coinsurance, and covered services). The survey also collected information on the establishment's tenure, type of ownership, industry, and location.

Major Limitations

Due to confidentiality restrictions, this survey is available to researchers only through the National Center for Health Statistics (NCHS) Research Data Center (see NCHS, 2007). Prospective researchers must submit research proposals to RDC. In addition, there are costs associated with working in the center (\$500 per month for remote access and \$1,000 per week of on-site access).

MEDICAL EXPENDITURES PANEL SURVEY

Data Collection Method and Coverage

The Medical Expenditures Panel Survey (MEPS), conducted by the Agency for Healthcare Research and Quality, also contains information on the health insurance offerings of businesses of different employment sizes. The MEPS includes four components that provide an important overview of access to health insurance and care. It includes a Household Component (HC), Nursing Home Component (NHC), Medical Provider Component (MPC), and Insurance Component (IC). HC and IC provide some information valuable to small business researchers. The MEPS collects data on health services used by Americans, frequency of use, cost of services, and method of payment. In addition, data is collected on the cost, scope, and breadth of private health insurance held by and available to the U.S. population.³⁴ Yearly surveys have been conducted since 1996.

The HC collects data from a sample of families and individuals around the nation. The sample is drawn from the nationally representative subsample of households that participated in the prior year's NEHIS. The households are interviewed for five rounds that cover two full calendar years.

The IC derives information from two subcomponents: the household sample and the list sample. The household sample includes employers and other insurance providers (unions and insurance companies) of respondents to the previous year's MEPS HC. The number of respondents in this sample varies from year to year. These data, when linked back to the original household respondent, allow for the analysis of individual

³⁴ See description of the program at U.S. Department of Health and Human Services (undated).

behavior and choices made with respect to health care use and spending.³⁵ The list sample includes business establishments and governments. The sampling frame is derived from the BR and the target size of the list sample is approximately 30,000 per year.

Main Variables

Firm and Establishment Size Variables. The HC survey collects data on the size of current or former (if not currently employed) employers at the firm level. The IC also collects information on a firm's total employment and employment at the chosen establishment.

Other Variables. The HC survey includes additional employment information such as type of industry, wage level, weekly hours of work, and current employment status (working, unemployed, or retired). In addition, it tracks changes in respondents' health status, income, employment, eligibility for public and private insurance coverage, use of services, and payment for care.

The IC also includes data on the number of employees eligible for coverage, employment characteristics, and full address. In addition, it collects information on the number of health insurance plans that were offered by the establishment, as well as other establishment characteristics.

Major Limitations

There are several limitations to using this data set. The questionnaire and coverage have changed somewhat over time. For example, information from the sample of self-employed workers was collected only in 1996 and was discontinued after that. New questionnaires were added in 1997 for cases in which the collection took place at company headquarters for 10 or more establishments. In 1998, additional questionnaires were added to deal with follow-up interviews for multiunit establishments at the company level.

In addition, due to confidentiality constraints, MEPS microdata can be accessed only through the Data Centers at the Agency for Healthcare Research and Quality.

What Can Be Done with the Data

Several studies have used MEPS. For example, Gresenz, Rogowski, and Escarce (2004) used the data to study access to care among the uninsured. This data set can also potentially be used to study characteristics of firms that provide health insurance coverage for workers. For example, Zawacki and Taylor (2005) use MEPS-IC to examine characteristics of establishments that paid 100 percent of health insurance premiums for their workers. In addition, the data set can be used to study responses of workers to different types of insurance provided by the firm (e.g., the selection of workers into different firms based on the health insurance coverage or compensation for workers who secure health insurance from other sources). Some of the component data sets allow for analysis of the self-employed, costs of providing insurance for small firms,

³⁵ See description at AHRQ (2007).

and the impact of changes in federal and state health care policies. Within the census RDCs, the MEPS data can be matched to the records obtained from the EC. For discussion of linking algorithms and possible issues that arise, see McCue and Zawacki (2005).

Chapter Five

PRIVATE AND COMMERCIALY AVAILABLE DATA SOURCES

Although government administrative data sets and surveys provide a wealth of information on small businesses, there are two important reasons for also considering private data sources. The first reason is that few of the government data sources are publicly available, including the longitudinal data sources, or the potential data frames of all businesses. We note that, while this is a substantial road block for researchers, it is for good reason, as much of the government data is collected on the condition of confidentiality. The second reason is that, as with all data sources, the information collected by government sources typically seeks to address a particular question or mandate and may not include the information necessary to address other timely policy questions on particular populations of interest.

The private data sources described here are publicly available but may carry considerable cost. In addition, the needs of researchers may not have been the primary concern for those collecting the data. Thus, these sources can raise a variety of other data issues such as coverage, representation, and, for survey sample sources, response rates.

DUN'S MARKET IDENTIFIER (DMI)

Data Collection Method and Coverage

The private data source most widely used by both government and private organizations for research on small businesses is Dun and Bradstreet's (D&B) list of U.S. businesses, the Dun's Market Identifier (DMI) file. While the Census Bureau and BLS use separate master files of U.S. businesses as sampling frames for their own surveys, they do not make these lists available to other government agencies or private organizations due to confidentiality concerns. For these agencies and organizations, the DMI is currently the most complete listing of businesses from which to potentially draw a nationally representative probability sample.

The DMI contains information on more than 14 million U.S. businesses. DMI data are continually updated based on information from a variety of sources, primarily D&B's credit rating service and business directories, but also from direct investigation and interviews, payment and banking data from company suppliers, suits, liens, judgments, business registrations, bankruptcy filings, corporate financial reports, government contracts, grants, loans and debarments, data mining of more than 27 million Internet domains, news and media sources, and print directories. The frequency of information updates for a business is a function of its employment, industry, and activity level.

Main Variables

Firm and Establishment Size Variables. The DMI includes measures of firm and establishment employment (employment includes owners or unpaid family members who are workers, which differs from the definition used by government data sources), and annual sales.

Other Variables. In addition, the listing provides detailed location data: telephone number, location, and owner name. Data are also collected for owner minority status, industrial classification, firm's start year, and legal status.

Major Limitations

The DMI is neither longitudinal nor does it provide a snapshot of information from a particular time point, since data are updated at different intervals for each business. It is difficult to gauge the range of DMI coverage, especially those of new and small businesses. Over time, D&B has expanded its coverage of small businesses and small establishments within larger firms. Several recent surveys that used the DMI as a sampling frame found it useful to first screen selected businesses to make sure that they were currently active and were private firms with employees. The NEHIS, for instance, found that approximately 18 percent of the screened businesses did not meet these criteria. The Kauffman Firm Survey (KFS, described below) will soon be able to provide more up-to-date information on the possible over- or undercoverage of the DMI.

What Can Be Done with the Data

The data is most commonly used as a sampling frame for new surveys of firms. DMI data are available for purchase from D&B, which makes the listing more accessible than other sources.

KAUFFMAN FIRM SURVEY

Data Collection Method and Coverage

The Kauffman Firm Survey (KFS) is a new survey commissioned by the Kauffman Foundation to provide publicly available longitudinal data on new firms. Researchers currently have completed two rounds of pilot data collection and the first full panel. Data from the baseline survey were projected to be available to researchers in the fall of 2006. The goal of the KFS is to longitudinally track new firms over time with an emphasis on financial development, high technology, and woman-owned firms. Two cohorts of new firms will each be followed for multiple years. The first cohort of approximately 5,000 firms, new in 2004, will be followed for three additional years and the second panel of 5,000 firms, new in 2006, for one additional year. The KFS is using listings from the DMI with a 2004 start year as the sampling frame for the first cohort and are oversampling high technology and woman-owned firms. Although details are not yet available, the expectation based on the pilot data is that only 40 percent of the firms selected will be eligible for the study due to errors in DMI data or overcoverage of the frame, and, of these, approximately 40 percent will complete the survey.

The first stage of sampling includes a 10-item screener for new firms. This screener includes questions about the timing of first-paying unemployment insurance taxes, payment of social security taxes, submission of a Schedule C for business income or losses, and application for an EIN. This information could be used to determine when new firms are recorded in government data sources (BITS and BED) that use one or more of these indicators to mark firm births. In addition, the screener collects information on possible forms of a firm's legal status (there are seven options including sole proprietorship, limited partnership, and limited liability company).

Main Variables

Firm and Establishment Size Variables. The full survey collects information about firm employment, including number of full- and part-time workers, payroll, and annual revenue. The employment count excludes owners and contract workers who are not on the firm's official payroll.

Other Variables. In addition, the survey collects information about the proprietor's work behavior and demographics, characteristics of the firm (including the number of patents, copyrights, and trademarks), business strategy and use of innovation, business organization, human resources benefits, and detailed information on finances.

Major Limitations

This survey is not yet available to researchers.

What Can Be Done with the Data

Once available, this data will allow researchers to track the experiences of new businesses. The oversampling of high technology and woman-owned firms makes the data especially relevant to researchers investigating the experiences of these types of firms. The data should be useful for understanding how the different indicators of the birth of a new firm are related to one another which will be useful for understanding the point at which new firms are captured in other data sources that rely on a particular indicator. The survey also includes a wealth of financial information which allows for comparison of financial tactics and decisions that are associated with success or failure and for a comparison of financial options, tactics, and decisions made by different types of businesses and business owners.

RESEARCH DATA SET DERIVED FROM THE MARTINDALE-HUBBELL LAW DIRECTORY

Data Collection Method and Coverage

In recent years, researchers have synthesized databases from various listings and directories. One example is a research database of law firms extracted from the Martindale-Hubbell Law Directory (MH). MH is the leading reference on the U.S. legal services industry. MH publishes listings for almost the entire

universe of lawyers and law firms, but the underlying data for each firm are not available. However, researchers can use directory listings to synthesize the database.

Romley and Talley (2005) extracted the listings of all lawyers and law firms from for 1993 and 1999. A matching mechanism was used to create a firm-level database with information on each establishment within a firm. There are approximately 65,000 law firms included each year and firms are connected between years.

Main Variables

Firm and Establishment Size Variables. The database includes office- and firm-level employment.

Other Variables. The database also includes information on the distribution of employment by type of position, number of offices within a firm, quality rankings, types of law fields in which the firm operates, and organizational form.

Major Limitations

There are some limitations to this constructed database. First, this research database is not publicly available, although researchers can design automated algorithms that would resynthesize the database. Second, the current matching algorithm is limited. The algorithm matches law offices to a law firm based on name only. This can lead to false matching of unrelated law firms that have the same name. This problem is especially pronounced in the case of small firms that use the lawyer's name as the firm name. The algorithm can be improved by also matching by year of the firm's inception. In addition, to date, this research database covers only two years. Additional years of the data can be extracted.

What Can Be Done with the Data

Romley and Talley (2005) used this longitudinal data set to examine effects of the availability of new organizational forms for law firms of different employment sizes. More generally, researchers can use this data set to examine the effects of different policies on access to and quality of lawyers as well as their specializations within the legal profession.

THE KAISER FAMILY FOUNDATION/HEALTH RESEARCH AND EDUCATIONAL TRUST EMPLOYER HEALTH BENEFITS SURVEYS

Data Collection Method and Coverage

There are also private databases with information about small firms and health insurance offerings. One such data source is the Kaiser Family Foundation/Health Research and Educational Trust (Kaiser/HRET) Employer Health Benefits Surveys, which have been fielded annually since 1999. Previous versions of the survey were sponsored by the Health Insurance Association of America from 1987 to 1990 and by KPMG Peat Marwick from 1991 to 1998. The nationally representative sample was drawn from the D&B

list of the nation's employers with three or more workers, stratified by firm employment and industry. Approximately 65 percent of the sample members in 2006 were participants in either the 2004 or 2005 survey or both, so some panel data are available. The survey covers about 2,000–3,000 employers each year.

Main Variables

Firm and Establishment Size Variables. The survey collects information on the number of workers in a firm. Firm size definitions are as follows: three to nine workers, 10 to 24 workers, 25 to 49 workers, 50 to 199 workers, 200 to 999 workers, 1,000 to 4,999 workers, and 5,000 or more workers.

Other Variables. The surveys collect detailed information on health insurance offerings, the cost of coverage, health management programs, and employer opinions about health insurance. Beginning in 2006, the survey collects full information on high deductible health plans with a savings option.

Major Limitations

Response rates for this survey in recent years have been 50 percent or lower and requests for data are handled on an individual basis. Another major limitation of this data set is that state identifiers are not publicly available, making it difficult to make cross-state comparisons.

What Can Be Done with the Data

Researchers used this data to examine effects of health insurance regulations on the firm size. For example, Kapur et al. (2006) used a repeated cross-section sample from 1993, 1996, and 1998 surveys to examine how small group health insurance reforms affected firm size.

CONCLUSION

Great strides have been made in recent years to create data sources useful for conducting research on small business and its policies. Of particular importance are new longitudinal data sets created by the Census Bureau and the Bureau of Labor Statistics, which allow for the study of business entry and exit (which is especially relevant to small business policy) as well as changes within establishments and firms to be studied over time.

In creating administrative longitudinal databases, progress has been made on three problematic issues: connecting establishments to parent firms, matching establishments and firms over time, and identifying firm inception dates and closures. This substantial work made it possible to create the data sets described above; however, challenges in using these data remain. At present, there continue to be significant challenges involved in obtaining longitudinal data, data that include information at several organizational levels, and appropriate sampling frames.³⁶ Two of the main problems are lack of availability due to cost or confidentiality concerns and the poor quality of linkages, either within units over time or between establishments and firms. Other important concerns for longitudinal data sources include the point in time at which a new firm or establishment is identified to enter a database, the point in time at which it is determined that a business has closed and should be removed, and the point in time at which the size of the business, whether based on number of employees or some other criteria, is measured.³⁷ Researchers need to carefully consider how well the three issues were addressed in each data source and whether the resulting quality of the data might impact their research.

Researchers interested in studying small business issues need to be aware of whether the size measures included in a particular data set reflect establishment size or firm size. Although some of the data sets described in this discussion allow for the linking of establishment and firms, many of them provide information on one but not the other. Although it is common for researchers to use establishment size as a proxy for firm size, the work of Mendeloff et al. (2006) illustrates that there are substantive differences between the two. Policy research on small business issues must be careful to determine the unit of analysis that is critical to the policy under investigation and ensure that the data are, in fact, capturing that unit of analysis.

Most of the data sets described in this paper are based on surveys of—or information on—businesses. Such data are useful for examining many different issues. The CPS, a household survey, provides a rich and widely used data resource for studying entrepreneurship and new business formation; however, it is not possible to study business growth, success, and failure with this data set. Firm-level data are needed to address

³⁶ We note that most government data sources are unlikely to be usable as sampling frames by researchers outside of the government due to the strong confidentiality provisions under which the data were obtained.

³⁷ For information on how size class measurement timing can matter, see Okolie (2004).

the latter issues. Although longitudinal firm-level data needed to address such issues are available from the Census Bureau, particularly the BITS and BED database, only a handful of research studies have used these sources. Indeed, officials at the Census Bureau have voiced interest in more researchers making use of these rich data sources. In his presentation at the Kauffman Symposium on Entrepreneurship Data (November, 2004), John Haltiwanger strongly encouraged researchers both to submit proposals for using the data at census RDCs and to continue to support the Census Bureau's use of the BED as a frame for survey samples through which more detailed and specific information can be collected and made publicly available.

The most notable gap in current small business data sources is the lack of a *publicly available* source of longitudinal data. In the next five years, this gap will be at least partially addressed by the Kauffman Firm Survey of new businesses. Information on this survey design and instrument is available now and researchers can begin to design research studies that would take advantage of the data when they become available.

Appendix A

KEY INFORMATION SOURCES FOR SMALL BUSINESS DATA

GENERAL RESOURCES ON DATA FOR SMALL BUSINESSES

History of Government Small Business Data Collection Efforts

Armington, Catherine, "Development of Business Data: Tracking Firm Counts, Growth, and Turnover by Size of Firms," *Small Business Research Summary*, No. 245, December 2004. As of June 2, 2006:
<http://www.sba.gov/advo/research/rs245tot.pdf>

Current Research and Resources from the SBA Office of Advocacy

U.S. Small Business Administration, *SBA Office of Advocacy, Office of Economic Research: Research Publications 2004*, Washington, D.C.: Small Business Administration, January 2005. As of July 5, 2005:
http://www.sba.gov/advo/research/res_pub04.pdf

U.S. Small Business Administration, *The Small Business Economy: A Report to the President*, Washington, D.C.: SBA Office of Advocacy, 2004. As of June 2, 2006:
http://www.sba.gov/advo/stats/sb_econ2004.pdf

SBA List of Links to Resources on Small Business Data

U.S. Small Business Administration, "FAQs: Frequently Asked Questions," undated Web page. This is a good, fairly complete list of government data sources on small business. As of July 5, 2005:
<http://app1.sba.gov/faqs/faqindex.cfm?areaID=2>

CENSUS BUREAU DATA SOURCES

Research Using the BITS

Acs, Zoltan J., and Catherine Armington, "Using Census BITS to Explore Entrepreneurship, Geography, and Economic Growth," *Small Business Research Summary*, No. 248, February 2005. As of July 5, 2005:
<http://www.sba.gov/advo/research/rs248tot.pdf>

National Women's Business Council, "Firms Owned by Women of Color Show Staying Power: African American Firms Lag Somewhat Behind," news release, Washington, D.C., August 24, 2004. As of June 2, 2006:
<http://www.nwbc.gov/documents/FINAL-Census-Bureau-Trends08.24.04.pdf>

National Women's Business Council, "Women Employer Firms Continue to Show Strength: Similar Survival Rates, Fewer Job Losses from 1997–2001," news release, Washington, D.C., February 14, 2005. As of June 2, 2006:
http://www.nwbc.gov/NewsCenter/documents/final_census_bureau_trends_news_release_02-09-05.pdf

U.S. Small Business Administration, *New Data for Dynamic Analysis: The Business Information Tracking Series (BITS)*, Washington, D.C.: U.S. Small Business Administration, Office of Advocacy, 2000. As of March 1,

2007:
<http://purl.access.gpo.gov/GPO/LPS4810>

Venegas, Ernesto C., “Issue Brief: Start-Ups and Expansions Fuel Minnesota’s Economic Engine,” St. Paul, Minn.: Minnesota Department of Trade and Economic Development, December 2000. As of June 2, 2006:
<http://www.deed.state.mn.us/bizdev/PDFs/startups&expan.pdf>

Villalonga, Belén, “Diversification Discount or Premium? New Evidence from BITS Establishment-Level Data,” *Journal of Finance*, Vol. 59, No. 4, April 2004, pp. 479–506. As of June 2, 2006:
<http://www.people.hbs.edu/bvillalonga/VillalongaJF2004.pdf>

Information on the 2002 EC

U.S. Census Bureau, “Economic Census,” 2002 reports, data tables, links to other economic census years. As of June 2, 2006:
<http://www.census.gov/econ/census02/>

Information on SWOBE, SMOBE, and SBO

U.S. Census Bureau, *2002 Survey of Business Owners: Advance Report on Characteristics of Employer Business Owners*, Washington, D.C.: U.S. Bureau of the Census, 2002. As of July 5, 2005:
<http://www.census.gov/econ/census02/sbo/intro.htm>

U.S. Census Bureau, *1997 Economic Census Survey of Minority- and Women-Owned Business Enterprises*, Washington, D.C.: U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2002. Links to surveys, reports, data. As of June 2, 2006:
<http://www.census.gov/csd/mwb/>

Information on the COS

U.S. Census Bureau, “Company Organization Survey,” last revised July 27, 2006. Overview and brief methodology description, links to CBP. As of June 2, 2006:
<http://www.census.gov/econ/overview/mu0700.html>

BUREAU OF LABOR STATISTICS DATA SOURCES

Research Using the BED Data

Spletzer, James R., R. Jason Faberman, Akbar Sadeghi, David M. Talan, and Richard L. Clayton, “Business Employment Dynamics: New Data on Gross Job Gains and Losses,” *Monthly Labor Review*, April 2004, pp. 29–42. As of June 2, 2006:
<http://www.bls.gov/opub/mlr/2004/04/art3full.pdf>

U.S. Department of Labor, Bureau of Labor Statistics, “Business Employment Dynamics,” undated Web page. Links to statistical resources and publications on business employment dynamics. As of June 2, 2006:
<http://www.bls.gov/bdm/home.htm#publications>

U.S. Department of Labor, Bureau of Labor Statistics, “State Publications Using Geocoded QCEW Data,” last modified August 30, 2004. Links to state Web sites containing geocoded QCEW research. As of June

2, 2006:
<http://www.bls.gov/bdm/geocode.htm>

Research Using the CES

U.S. Department of Labor, Bureau of Labor Statistics, “Employment, Hours, and Earnings from the Current Employment Statistics Survey (National),” Undated Web page. As of June 2, 2006:
<http://www.bls.gov/ces/home.htm>

Research on Small Businesses Using the CPS

SBA Office of Advocacy, “Characteristics of Small Business Employees and Owners, 1997,” January 1998. A reference guide on the workers and owners of small businesses prepared by the Office of Economic Research of the U.S. Small Business Administration’s Office of Advocacy. As of June 2, 2006:
http://www.sba.gov/ADVO/stats/ch_emp_o.html

Data on Nonemployer and Employer Businesses—the ILBD

Davis, Steven J., John Haltiwanger, Ron S. Jarmin, C. J. Krizan, Javier Miranda, Alfred Nucci, and Kristin Sandusky, *Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes*, Washington, D.C.: Bureau of the Census, 2006. As of June 1, 2006:
http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101744

OTHER GOVERNMENT DATA SOURCES

Information on the SSBF Including Methodology Reports, Papers, Codebooks, Survey Instruments, and Data

Federal Reserve Board, Survey of Small Business Finances, “Occasional Staff Studies,” last updated November 20, 2006. Links to working papers and methodology reports, codebooks and other related documentation, and the full public data sets are available for the 1998, 1993, and 1987 SSBF. A list of answers to FAQs is also available. As of June 2, 2006:
<http://www.federalreserve.gov/pubs/oss/oss3/nssbftoc.htm>

Information on the IRS Survey of Income-Related Research Projects

Internal Revenue Service, “SOI Projects and Contacts,” undated. Survey of income, projects, and contacts. As of June 2, 2006:
<http://www.irs.gov/pub/irs-soi/projcont.pdf>

Information on the Survey of Small Businesses’ Telecommunications Use and Spending for SBA.

Pociask, Stephen B., *A Survey of Small Businesses’ Telecommunications Use and Spending*, Washington, D.C.: Small Business Administration, March 2004. As of June 2, 2006:
<http://www.sba.gov/advo/research/rs236tot.pdf>

PRIVATE SOURCES OF DATA

D&B

Dun and Bradstreet, “The DUNSRight™ Process: The Power Behind Quality Information,” undated Web page. Description of the system used to ensure data quality. As of July 1, 2005:
http://www.dnb.com/us/about/db_database/dnbinfoquality.html

Standard and Poor’s Compustat Data Resources for Research

Standard and Poor’s, “Compustat,” undated Web page. Description of software and ordering information. As of June 2, 2006:
http://www.compustat.com/www/ug/univ_mkt.html

Information on the 2004 Kaiser/HRET Employer Health Benefits Survey

Henry J. Kaiser Family Foundation, and Health Research and Educational Trust, *Employer Health Benefits: 2004 Annual Survey*, Menlo Park, Calif.: Henry J. Kaiser Family Foundation, 2004. This annual survey of employers provides detailed insights into trends in employer-based health coverage, including changes in premiums, employee contributions, cost-sharing policies and other relevant information. As of June 2, 2006:
<http://www.kff.org/insurance/7148/index.cfm>

The Center for Research in Security Prices (CRSP) Data Resources for Research

Center for Research in Security Prices, undated homepage. CRSP data include firm-level financial and stock price information, the data can be linked to Compustat using CUSIP® identifiers.³⁸ As of June 2, 2006:
<http://www.crsp.chicagogsb.edu/products/index.html>

Securities Data Company (SDC) Resources

Fisher College of Business, “Securities Data Company,” last updated August 14, 1995. SDC data has been used to analyze the impact of the Sarbanes-Oxley Act on firm decisions to exit the public securities market. As of June 2, 2006:
http://fisher.osu.edu/fin/resources_data/provider/sdc.htm

CONFERENCE RESOURCES

Entrepreneurship in the 21st Century (March 26, 2004), Conference Proceedings

U.S. Small Business Administration, and the Ewing Marion Kauffman Foundation, *Entrepreneurship in the 21st Century*, conference proceedings, March 26, 2004. As of June 2, 2006:
http://www.sba.gov/advo/stats/proceedings_a.pdf

U.S. Small Business Administration, and the Ewing Marion Kauffman Foundation, *Summaries of the Presentations*, conference proceedings, March 26, 2004. As of June 2, 2006:
http://www.sba.gov/advo/stats/proceedings_b.pdf

U.S. Small Business Administration, and the Ewing Marion Kauffman Foundation, *Appendix D: PowerPoint Presentations*, conference proceedings, March 26, 2004. As of June 2, 2006:
http://www.sba.gov/advo/stats/proceedings_c.pdf

³⁸ CUSIP® is a registered trademark of the American Banker’s Association. CUSIP stands for Committee on Uniform Security Identification Procedures.

Kauffman Symposium on Entrepreneurship Data (November 10–11, 2004)

Conference Summary Document available by request from Amelia Haviland

NATIONAL ACADEMY OF SCIENCES PANELS

National Academies, “Measuring Business Formation, Dynamics and Performance,” undated Web page.

Links to project scope and panel membership. As of June 2, 2006:

<http://www8.nationalacademies.org/cp/projectview.aspx?key=92>

Review of Federal Business Statistics, National Academies, Division of Behavioral and Social Sciences Education, Committee on National Statistics. Link to latest meeting agenda

<http://www8.nationalacademies.org/cp/projectview.aspx?key=CNST-I-04-01-A> (accessed June 2, 2006).

Panel Members: John Haltiwanger (co-chair), Lisa Lynch (co-chair), John Abowd, Patricia Anderson, Matthew Barnes, Steven Davis, Timothy Dunne, Robert Groves, Susan Hanson, Robert McGuckin, Paul Reynolds, Mark Roberts, Niels Westergard-Nielsen, Kirk Wolter. Final Report will be issued January 2006.

Appendix B

SUMMARY OF THE AVAILABLE DATA SETS

DESCRIPTION OF THE MAIN PARTS OF THE TABLE

Collection method and coverage contains information about the origin and sources of the data. Note that some of the entries refer to tables aggregated from the actual data set. The number of observations includes the most recent date for which the data are reported unless stated otherwise.

Topic, main variables contains establishment or firm size variables (shown in bold type) in addition to the most important variables in the data set and information on the most precise geographical identifiers available (region, state, county, ZIP code, city, full address).

Periodicity and dates available contains the range of years for which data are available. This column also indicates whether the data are cross sectional or whether the underlying units of observations are connected over time into a panel.

Unit of observation contains units for which data are gathered and any linkages to higher levels of observation (e.g., establishment-level data aggregated at the level of firm).

Employment definition is typically defined as the number of people on payroll in the pay period that includes the 12th of the month.

Major limitations contains the main limitations of the data set.

Table A.1
Summary of the Available Data Sets

Name of the Data Set; Source	Collection Method and Coverage	Topic, Main Variables	Periodicity, Dates Available, Longitudinal Links	Unit of Observation	How Employee Is Defined	Major Limitations
QCEW, also known as ES-202; BLS	Administrative records. Includes all establishments covered by UI and UCFE, about 8.4 million establishments	Employment, wages, full address (both mailing and physical location)	Quarterly, cross section; 2001 forward (NAICS basis); 1975–2000 (SIC basis)	Establishment, can be aggregated to the firm level using EIN	Everyone on payroll	Not publicly available. ^a Some aggregated tables are available. ^b Excludes self-employed, unpaid family members, elected officials. UI coverage is different by state.
BED; BLS	Connects 6.4 million establishments from QCEW over time using SESA-ID and probability matching	Monthly employment, wages, job gains and losses, full address	Quarterly, panel, 1992–forward	Establishment, can be aggregated to the firm level using EIN	Everyone on payroll	Not publicly available. ^a Some aggregated tables are available. Excludes government employees, private households, and establishments with zero employment. UI coverage differs by state and may change over time.
CES; BLS	Monthly sample survey of about 160,000 businesses and government agencies covering about 400,000 establishments	Employment, hours, and earnings, industry detail, full address	Monthly, cross section, 1990–ongoing, some series are available since 1939	Establishment	Everyone on payroll	Not publicly available. ^a Some aggregated tables are available. ^d Establishments are not connected over time. Nonresponse.
NCS; BLS	Survey, sampling frame is QCEW, three-stage design: regions, establishments, and occupations. About 19,000 establishments. Large firms are more likely to be selected.	Benefits and wages, firm employment	Yearly, cross section	Occupations within an establishment	Everyone on payroll	Not publicly available. ^a Change in methodology in 1999.

^aResearcher can apply for the access to the confidential microdata. For details, see BLS (2007b).

^bSee, for example, BLS (undated[c]).

^cSee, for example, BLS (2007c).

^dSee, for example, BLS (undated[a]).

Name of the Data Set; Source	Collection Method and Coverage	Topic, Main Variables	Periodicity, Dates Available, Longitudinal Links	Unit of Observation	How Employee Is Defined	Major Limitations
---------------------------------	--------------------------------	-----------------------	--	------------------------	-------------------------------	-------------------

^oResearchers can apply for access to confidential data (U.S. Bureau of the Census, undated[b]).

Name of the Data Set; Source	Collection Method and Coverage	Topic, Main Variables	Periodicity, Dates Available, Longitudinal Links	Unit of Observation	How Employee Is Defined	Major Limitations
CPS; BLS, Census Bureau	Monthly sample survey of approximately 60,000 households. Rotating sample design, respondents are in for 4 months, out for 8 months, and in for an additional 4 months.	Firm employment, business ownership, self-employment, some characteristics of small business employee	Monthly, 1962–ongoing. Respondents are in for 4 months, out for 8 months, and in for an additional 4 months.	Household, family, person	Not explicitly discussed	Categorical size count, matching over time is imperfect
SSEL or BR; Census Bureau	List of all establishments and companies with paid employees; 180,000 multiunit companies, representing 1.5 million affiliated establishments, 5 million single-establishment companies, and nearly 14 million nonemployer businesses. Administrative data from IRS and SSA. Also compiles data from economic censuses and current business surveys.	Employment, revenues, business full address, organization type, industry classification, operating data, EIN.	Yearly, cross section, 1974–2001	Establishment and firm	N/A	Not publicly available.
LBD; Census Bureau	Matched records from SSEL over time using PPN, using CFN and EIN, or using name and address match. Covers all nonfarm private economy and some public sector activities. 4.5–7.1 million records per year.	Establishment age and tenure, payroll, employment, firm affiliation, full address	Yearly, panel, 1974–1999, ongoing	Establishment, firm	Everyone on payroll	Not publicly available. ^a
BITS, also known as LEEM; Census Bureau	Links SSEL establishments over time using PPN, CFN, or EIN. Includes establishments with positive payroll. 13 million establishments. Same industry coverage as CBP.	Employment, firm employment, payroll, firm ownership, firm affiliation, census geography, primary industry, starting year, census file number	Yearly, panel, 1989–ongoing	Establishment, firm	Everyone on payroll	Not publicly available. ^a

^aResearchers can apply for access to confidential data (U.S. Bureau of the Census, undated[b]).

Name of the Data Set; Source	Collection Method and Coverage	Topic, Main Variables	Periodicity, Dates Available, Longitudinal Links	Unit of Observation	How Employee Is Defined	Major Limitations
CBP; Census Bureau	Aggregated tables derived from SSEL, excludes some agriculture, rail transportation; private households; and public administration.	Employment, payroll , total number of establishments, county	Yearly, cross section, 1977–forward	Establishment, firm	Everyone on payroll	Aggregated tables, some industry level data is not disclosed
ILBD; Census Bureau	Connects establishment data from LBD to statistics of nonemployers	See LBD	Yearly, panel, 1992–2001	Establishment	Everyone on payroll	Not publicly available. ^a
LEHD; Census Bureau	Connects establishment data from LBD to household data. 4 million establishments for about 20 states.	See LBD and in addition: employer human capital, workforce indicators	Yearly, panel, 2003	Establishment	Everyone on payroll	Not publicly available. ^a
EC; Census Bureau	Covers 5 million establishments with more than five employees and a sample of the rest.	Employment , labor costs, measures of output, expenses, city identifiers	Years ending in 2 and 7	Establishment, firm	Everyone on payroll	Not publicly available. ^a
COS), also known as Report of Organization; Census Bureau	Surveys 40,000 multiunit companies with more than 250 employees, and approximately 10,000 smaller multiunit companies on rotating basis.	Establishment operational status, payroll, employment , controlling interests held by other domestic or foreign-owned organizations	Annually since 1974, cross section, survey coverage and content vary during the census year	Establishment, firm	Everyone on payroll	Not publicly available. ^a
SWOBE, SMOBE; Census Bureau	Sample from BITS data frame, part of EC	Organizational form, sales and receipts, employees and annual payroll .	1992, 1997, 2002, cross section	Establishment	Everyone on payroll	Not publicly available. ^a
CBO; Census Bureau	Sample from BITS data frame, 78,000–115,000 records for establishments and 117,000–128,000 observations for owners file, part of EC	Legal form of organization, receipts , sources of capital, employment , whether the business is home based or not.	Yearly, 1982, 1987, 1992; combined with the SMOBE and SWOBE in 2002 to form SBO, cross	Establishment and individuals	All employees reported on a firm's payroll during specified pay periods	Not publicly available. ^a

^aResearchers can apply for access to confidential data (U.S. Bureau of the Census, undated[b]).

SBO; Census Bureau	Sample from BITS data frame, 78,000–115,000 records	Legal form of organization, receipts , sources of capital, employment , whether the business is home based or not.	Yearly, cross section, 2002; for other years, see SMOBE and SWOBE	Establishment	All employees on payroll	Not publicly available. ^a
SOI; IRS	Stratified probability samples of master file of all tax returns	Tax related issues, business receipts, selected deductions, payroll, and net income	Yearly, cross section, 1990–2002	Firm	All employees on payroll	Not publicly available.
SSBF; Federal Reserve Board	Sampling frame is the DMI file. Firms with fewer than 500 employees. About 3,500 businesses.	Firm’s use of credit, firm’s assets, liabilities, income, revenues , profits, expenses, employment , owners’ characteristics	1987, 1993, 1998, cross section	Firm	Employee on payroll or not, family members on payroll	Only 33% response rate.
NEHS ; Centers for Disease Control and Prevention	National probability sample survey of business establishments, governments, and self-employed individuals with no employees and no other locations. 34,604 completed interviews (70% response rate).	Health insurance offerings, employment	1994, cross section	Establishment	All employee on payroll	Not publicly available. ^b
Research data set derived from MH	Directory of lawyers and law firms. Complex algorithms can be used to extract and match records of lawyers to establishments and firms.	Firm and office employment , specialization, ratings, full address	1993, 1999, panel	Law firm, law office	Lawyers or supporting personnel affiliated with the firm	Not publicly available

^bResearchers can apply for access to confidential data (U.S. Bureau of the Census, undated[b]).

^aResearchers can apply for access to confidential data (NCHS, 2007).

Name of the Data Set; Source	Collection Method and Coverage	Topic, Main Variables	Periodicity, Dates Available, Longitudinal Links	Unit of Observation	How Employee Is Defined	Major Limitations
Kaiser/HRET Employer Health Benefits Surveys; Kaiser Family Foundation and Research and Educational Trust	Survey of public and private employers, sampled from DMI; about 3,262	Employer health plans coverage, costs, enrollment patterns, health plan choice, and employee costs, employment	Annually since 1999; before that, the survey was conducted by KPMG from 1991–1998, and by Health Insurance Association from 1987 until 1991.	Establishment	Not explicitly discussed	Not publicly available. Categorical definition of firm size.
MEPS; Agency for Healthcare Research and Quality	HC: Panels of 5 rounds of interviews over 30 months. IC: annual survey of establishments from SSEL sample frame, 27K establishments. Also sample of establishments with workers from prior year HC.	HC: Health status, access to care, income, employment , employment status, eligibility for private and public insurance coverage, health care use and expenses. IC: types of plans provided, number of workers covered, employment : total, by gender, by age over 50, and by earnings.	Annual, 1996–ongoing, HC: panel, IC: cross section	HC: household, IC: establishment	Everyone on payroll, does not include temporary workers	Not publicly available. ^a
DMI; D&B	Extension of D&B credit database	Info about owners, sales, employment and legal status, full address	Yearly, ongoing, panel can be created using D&B identifiers	Establishment, firm	Includes owners or unpaid family members	Is available for a fee
KFS; Kauffman Foundation	Survey of new businesses from DMI listing sample frame. About 5000 firms.	Owner's characteristics, employment , business organization and benefits, and business finances	Annual, 2005, panel	Owner		Will be publicly available, data have not yet been released.

³Researchers can apply for access to confidential data (NCHS, 2007).

REFERENCES

- Abowd, John M., John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, and Kristin Sandusky, *The Relation Among Human Capital, Productivity and Market Value: Building Up from Micro Evidence*, Washington, D.C.: U.S. Census Bureau, December 6, 2002. As of March 1, 2007:
http://instruct1.cit.cornell.edu/~jma7/nbercriw_20021206.pdf
- Abowd, John, John Haltiwanger, Julia Lane, and Kristin Sandusky, *Within and Between Firm Changes in Human Capital, Technology, and Productivity*, Washington, D.C.: U.S. Census Bureau, October 6, 2001. As of March 1, 2007:
<http://lehd.dsd.census.gov/led/library/misc/hcpaper.pdf>
- Acs, Zoltan J., and Catherine Armington, *Longitudinal Establishment and Enterprise Microdata (LEEM) Documentation*, Washington, D.C.: U.S. Bureau of the Census, May 14, 1998. As of March 1, 2007:
http://www.ces.census.gov/index.php/ces/1.00/cespapers?down_key=100265
- , “Job Creation and Persistence in Service and Manufacturing,” Discussion Papers on Entrepreneurship, Growth and Public Policy, 2004-16, Max Planck Institute of Economics, Group for Entrepreneurship, Growth, and Public Policy, 2004.
- , *Job Flow Dynamics in the Service Sector*, Washington, D.C.: Center for Economic Studies, U.S. Dept. of Commerce, Bureau of the Census, 1999. As of March 1, 2007:
<http://www.ces.census.gov/ces.php/abstract?paper=101581>
- , “Using Census BITS to Explore Entrepreneurship, Geography, and Economic Growth,” *Small Business Research Summary*, No. 248, February 2005. As of July 5, 2005:
<http://www.sba.gov/advo/research/rs248tot.pdf>
- Acs, Zoltan J., Catherine Armington, and Alicia Robb, “Measures of Job Flow Dynamics in the U.S. Economy,” Washington, D.C.: U.S. Small Business Administration, January 1999. As of June 1, 2006:
<http://www.sba.gov/advo/research/rs192tot.pdf>
- Agency for Healthcare Research and Quality, “Medical Expenditure Panel Survey.” As of June 15, 2007:
<http://www.meps.ahrq.gov/mepsweb/>
- Agesa, Jacqueline, Richard U. Agesa, and Gary A. Hoover, “Market Structure and Racial Earnings: Evidence from Job-Changers,” *The American Economic Review*, Vol. 91, No. 2, May 2001, pp. 169–173.
- Antos, Joseph R., “Union Effects on White-Collar Compensation,” *Industrial and Labor Relations Review*, Vol. 36, No. 3, April 1983, pp. 461–479.
- Armington, Catherine, and Zoltan J. Acs, “Job Creation and Persistence in Services and Manufacturing,” *Journal of Evolutionary Economics*, Vol. 14, No. 3, July 2004, pp. 309–325.

- Barsky, Carl B., “Incidence Benefits Measures in the National Compensation Survey,” *Monthly Labor Review*, Vol. 127, No. 8, August 2004, pp. 21–28. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2004/08/art3full.pdf>
- Bates, Timothy, *Small Businesses Do Appear to Benefit from State-Local Government Economic Development Assistance*, Washington, D.C.: Bureau of the Census, 1995. As of June 1, 2006:
http://www.webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=100238
- Becker, Randy, Joel Elvery, Lucia Foster, C. J. Krizan, Sang Nguyen, and David Talan, *A Comparison of the Business Registers Used by the Bureau of Labor Statistics and the Bureau of the Census*, Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics, August 2005. As of March 1, 2007:
<http://www.bls.gov/ore/pdf/st050270.pdf>
- Berger, Mark C., *Distribution of Low-Wage Workers by Firm Size in the United States: Final Report*, Lexington, Ky.: Carolyn Loeff and Associates, 1999. As of March 1, 2007:
<http://www.sba.gov/advo/research/rs196tot.pdf>
- Berstein, Jared, and Maury Gittleman, “Exploring Low-Wage Labor with the National Compensation Survey,” *Monthly Labor Review*, Vol. 126, No. 11–12, November–December 2003, pp. 3–12. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2003/11/art1full.pdf>
- Bitler, Marianne, *Small Businesses and Computers: Adoption and Performance*, San Francisco, Calif.: Federal Reserve Bank of San Francisco, 2001. As of March 1, 2007:
<http://www.frbsf.org/publications/economics/papers/2001/wp01-15bk.pdf>
- Bitler, Marianne P., Alicia M. Robb, and John D. Wolken, “Financial Services Used by Small Businesses: Evidence from the 1998 Survey of Small Business Finances,” *Federal Reserve Bulletin*, Vol. 87, No. 4, April 2001, pp. 183–205. As of March 1, 2007:
<http://www.federalreserve.gov/pubs/bulletin/2001/0401lead.pdf>
- BLS—see U.S. Bureau of Labor Statistics.
- Boden, Richard J., “Analyses of Business Dissolution by Demographic Category of Business Ownership,” *Small Business Research Summary*, No. 204, March 2001. As of June 1, 2006:
<http://www.sba.gov/advo/research/rs204tot.pdf>
- Bowlus, Audra J., Nicholas M. Kiefer, and George R. Neumann, “Estimation of Equilibrium Wage Distributions with Heterogeneity,” *Journal of Applied Econometrics*, Vol. 10, special issue, December 1995, pp. S119–S131.
- Brown, Charles, and James Medoff, “The Employer Size-Wage Effect,” *The Journal of Political Economy*, Vol. 97, No. 5, October 1989, pp. 1027–1059.
- Brown, Clair, John C. Haltiwanger, and Julia I. Lane, *Economic Turbulence: Is a Volatile Economy Good for America?* Chicago, Ill.: University of Chicago Press, 2006.

- Butani, Shail J., Richard L. Clayton, Vinod Kapani, James R. Spletzer, David M. Talan, and George S. Werking, Jr., "Business Employment Dynamics: Tabulations by Employer Size," *Monthly Labor Review*, Vol. 129, No. 2, February 2006, pp. 3–22. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2006/02/art1full.pdf>
- Butani, Shail, George Stamas, and Michael Brick, "Sample Redesign for the Current Employment Statistics Survey," *Proceedings of the Survey Research Methods Section, American Statistical Association*, Alexandria, Va.: American Statistical Association, 1997, pp. 517–522. As of June 13, 2006:
http://www.amstat.org/sections/srms/proceedings/papers/1997_088.pdf
- Card, David, "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, Vol. 64, No. 4, July 1996, pp. 957–979.
- Carrington, William J., Kristin McCue, and Brooks Pierce, "The Role of Employer/Employee Interactions in Labor Market Cycles: Evidence from the Self-Employed," *Journal of Labor Economics*, Vol. 14, No. 4, October 1996, pp. 571–602.
- , "Using Establishment Size to Measure the Impact of Title VII and Affirmative Action," *The Journal of Human Resources*, Vol. 35, No. 3, Summer 2000, pp. 503–523.
- Coleman, Susan, "The Borrowing Experience of Black and Hispanic-Owned Small Firms: Evidence from the 1998 Survey of Small Business Finances," *Academy of Entrepreneurship Journal*, Vol. 8, No. 1, 2002a, pp. 1–20. As of March 1, 2007:
<http://www.alliedacademies.org/Publications/Journals/aej8-1.pdf>
- , "Characteristics and Borrowing Behavior of Small, Women-Owned Firms: Evidence from the 1998 Survey of Small Business Finances," *Journal of Entrepreneurial Finance and Business Ventures*, Vol. 14, No. 2, 2002b, pp. 151–166.
- , "Borrowing Patterns for Small Firms: A Comparison by Race and Ethnicity," *Journal of Entrepreneurial Finance and Business Ventures*, Vol. 7, No. 3, 2003, pp. 87–108.
- Copeland, Kennon R., "Nonresponse Adjustment in the Current Employment Statistics Survey," *2003 Federal Committee on Statistical Methodology Conference Papers*, 2003, pp. 80–90. As of July 5, 2005:
<http://www.fcsm.gov/03papers/Copeland.pdf>
- Cutler, David M., and Brigitte C. Madrian, "Labor Market Responses to Rising Health Insurance Costs: Evidence on Hours Worked," *The RAND Journal of Economics*, Vol. 29, No. 3, Autumn 1998, pp. 509–530.
- Davis, Steven J., John Haltiwanger, Ron S. Jarmin, C. J. Krizan, Javier Miranda, Alfred Nucci, and Kristin Sandusky, *Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes*, Washington, D.C.: Bureau of the Census, 2006. As of June 1, 2006:
http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101744

- Dietz, Elizabeth, "Trends in Employer-Provided Prescription-Drug Coverage," *Monthly Labor Review*, Vol. 127, No. 8, August 2004, pp. 37–45. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2004/08/art5full.pdf>
- Evans, David S., and Linda S. Leighton, "Some Empirical Aspects of Entrepreneurship," *The American Economic Review*, Vol. 79, No. 3, June 1989, pp. 519–535.
- Even, William E., and David A. Macpherson, "Employer Size and Labor Turnover: The Role of Pensions," *Industrial and Labor Relations Review*, Vol. 49, No. 4, July 1996, pp. 707–728.
- Faberman, R. Jason, "Job Creation and Destruction Within Washington and Baltimore," *Monthly Labor Review*, Vol. 124, No. 9, September 2001, pp. 24–31. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2001/09/art3full.pdf>
- , "Job Flows and Labor Dynamics in the U.S. Rust Belt," *Monthly Labor Review*, Vol. 125, No. 9, September 2002, pp. 3–10. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2002/09/art1full.pdf>
- , *Gross Job Flows Over the Past Two Business Cycles: Not All "Recoveries" Are Created Equal*, Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics, working paper 372, June 2004. As of June 1, 2006:
<http://www.bls.gov/ore/pdf/ec040020.pdf>
- Fairlie, Robert, "Technology and Entrepreneurship: A Cross-Industry Analysis of Access to Computers and Self-Employment," *Small Business Research Summary*, No. 259, June 2005. As of March 1, 2007:
<http://www.sba.gov/advo/research/rs259tot.pdf>
- Fairlie, Robert W., and Alicia M. Robb, *Families, Human Capital, and Small Business: Evidence from the Characteristics of Business Owners Survey*, Washington, D.C.: Bureau of the Census, 2005. As of June 1, 2006:
http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101720
- Federal Reserve Board, "Abstracts and Bibliography: Survey of Small Business Finances," last updated October 24, 2006. As of June 21, 2005:
<http://www.federalreserve.gov/pubs/oss/oss3/abstract.html>
- Foster, Lucia, *Establishment and Employment Dynamics in Appalachia: Evidence from the Longitudinal Business Database*, Washington, D.C.: U.S. Bureau of the Census, December 2003. As of August 25, 2005:
http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101690
- Garicano, Luis, and Thomas N. Hubbard, "Learning About the Nature of Production from Equilibrium Assignment Patterns," April 2005a. As of March 1, 2007:
<http://www.kellogg.northwestern.edu/faculty/hubbard/htm/research/papers/humancapital42005.pdf>
- , *Managerial Leverage Is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyers' Human Capital*, London: Centre for Economic Policy Research, 2005b.

- Gresenz, Carole Roan, Jeannette Rogowski, and Jose J. Escarce, *Health Care, Markets, the Safety Net and Access to Care Among the Uninsured*, Santa Monica, Calif.: RAND Corporation, WR-215, 2004. As of March 1, 2007:
http://www.rand.org/pubs/working_papers/WR215/
- Gruber, Jonathan, and James Poterba, "Tax Incentives and the Decision to Purchase Health Insurance: Evidence from the Self-Employed," *The Quarterly Journal of Economics*, Vol. 109, No. 3, August 1994, pp. 701–733.
- Hatch, Julie, "Employment in the Public Sector: Two Recessions' Impact on Jobs," *Monthly Labor Review*, Vol. 127, No. 10, October 2004, pp. 38–47. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2004/10/art3full.pdf>
- Headd, Brian, "The Characteristics of Business Owners Database, 1992," Working Papers 99-8, Center for Economic Studies, U.S. Census Bureau, 1999.
- Headd, Brian, *Business Success: Factors Leading to Surviving and Closing Successfully*, Washington, D.C.: Center for Economic Studies, U.S. Dept. of Commerce, Bureau of the Census, 2001. As of March 1, 2007:
<http://www.ces.census.gov/ces.php/abstract?paper=101614>
- Helwege, Jean, and Frank Packer, "The Decision to Go Public: Evidence from Mandatory SEC Filings of Private Firms," Columbus, Ohio: Fisher College of Business, Ohio State University, 2003.
- Hetrick, Ron L., "Analyzing the Recent Upward Surge in Overtime Hours," *Monthly Labor Review*, Vol. 123, No. 2, February 2000, pp. 30–33. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2000/02/art3full.pdf>
- Hirsch, Barry T., David A. Macpherson, and J. Michael DuMond, "Workers' Compensation Reciprocity in Union and Nonunion Workplaces," *Industrial and Labor Relations Review*, Vol. 50, No. 2, January 1997, pp. 213–236.
- Hirsch, Barry T., and Edward J. Schumacher, "Unions, Wages, and Skills," *The Journal of Human Resources*, Vol. 33, No. 1, Winter 1998, pp. 201–219.
- Holmes, Thomas J., and James Andrew Schmitz, *On the Turnover of Business Firms and Business Managers*, Washington, D.C.: Bureau of the Census, 1992.
- Idson, Todd L., and Daniel L. Feaster, "A Selectivity Model of Employer-Size Wage Differentials," *Journal of Labor Economics*, Vol. 8, No. 1, January 1990, pp. 99–122.
- Jarmin, Ron S., Shawn D. Klimek, and Javier Miranda, *Firm Entry and Exit in the U.S. Retail Sector, 1977–1997*, Washington, D.C.: U.S. Census Bureau, 2004. As of March 1, 2007:
http://www.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101704
- Jarmin, Ronald S., and Javier Miranda, *The Longitudinal Business Database*, Washington, D.C.: Center for Economic Studies, U.S. Census Bureau, 2002. As of June 20, 2005:
<http://www.ces.census.gov/paper.php?paper=101647>

- Kapur, Kanika, Pinar Karaca-Mandic, Susan M. Gates, and Brent D. Fulton, *Do Small Group Health Insurance Regulations Influence Small Business Growth?* Santa Monica, Calif.: RAND Corporation, WR-351-ICJ, 2006. As of March 1, 2007:
http://www.rand.org/pubs/working_papers/WR351/
- Karoly, Lynn A., and Julie Zissimopoulos, *Self-Employment Trends and Patterns Among Older U.S. Workers*, Santa Monica, Calif.: RAND Corporation, WR-136, 2004. As of March 1, 2007:
http://www.rand.org/pubs/working_papers/WR136/
- Kirkland, Katie, “On the Decline in Average Weekly Hours Worked,” *Monthly Labor Review*, Vol. 123, No. 7, July 2000, pp. 26–31. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2000/07/art3full.pdf>
- Konigsberg, Sheryl, David Talan, and Richard Clayton, *The Geospatial Distribution of Employment: Examples from the Bureau of Labor Statistics Quarterly Census of Employment and Wages Program*, 2005. As of May 31 2006:
http://www.fcs.m.gov/05papers/Konigsberg_Talan_Clayton_VC.pdf
- Krantz, Rachel, “Employment in Business Services: A Year of Unprecedented Decline,” *Monthly Labor Review*, Vol. 125, No. 4, April 2002, pp. 17–24. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2002/04/art2full.pdf>
- Kropf, Jurgen, and Patricia Getz, “Noneconomic Fluctuations in Hours and Earnings Data,” *Monthly Labor Review*, Vol. 122, No. 8, August 1999, pp. 20–27. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/1999/08/art4full.pdf>
- Kuebler, Caryn, and Christopher D. Mackie, *Improving Business Statistics Through Interagency Data Sharing: Summary of a Workshop*, Washington, D.C.: National Academies Press, 2006.
- LEHD Program, *The Longitudinal Employer-Household Dynamics Program: Employment Dynamics Estimates Project Versions 2.2 and 2.3*, Suitland, Md.: U.S. Bureau of the Census, 2002. As of June 21, 2005:
<http://lehd.dsd.census.gov/led/library/techpapers/tp-2002-05-rev1.pdf>
- Lel, Ugur, and Gregory F. Udell, “Financial Constraints, Start-Up Firms and Personal Commitments,” Kelley School of Business Working Paper, Indiana University, 2002.
- Loewenstein, Mark A., and James R. Spletzer, “Delayed Formal On-the-Job Training,” *Industrial and Labor Relations Review*, Vol. 51, No. 1, October 1997, pp. 82–99.
- Macpherson, David A., and Barry T. Hirsch, “Wages and Gender Composition: Why Do Women’s Jobs Pay Less?” *Journal of Labor Economics*, Vol. 13, No. 3, July 1995, pp. 426–471.
- Madrian, Brigitte C., and Lars Lefgren, *A Note on Longitudinally Matching Current Population Survey (CPS) Respondents*, Cambridge, Mass.: National Bureau of Economic Research, 1999. As of June 9, 2005:
<http://www.nber.org/papers/T0247>

- McCue, Kristin, and Alice M. Zawacki, *Using Census Business Data to Augment the MEPS-IC*, Washington, D.C.: Bureau of the Census, 2005. As of June 13, 2006:
http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101736
- Mellow, Wesley, “Employer Size and Wages,” *The Review of Economics and Statistics*, Vol. 64, No. 3, August 1982, pp. 495–501.
- Mendeloff, John, Christopher Nelson, Kilkon Ko, and Amelia Haviland, *Small Businesses and Workplace Fatality Risk: An Exploratory Analysis*, Santa Monica, Calif.: RAND Corporation, TR-371-ICJ, 2006. As of March 1, 2007:
http://www.rand.org/pubs/technical_reports/TR371/index.html
- Morisi, Teresa L., “Recent Changes in the National Current Employment Statistics Survey,” *Monthly Labor Review*, Vol. 126, No. 6, June 2003, pp. 3–13. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2003/06/art1full.pdf>
- National Center for Health Statistics, *Employer-Sponsored Health Insurance: State and National Estimates*, Hyattsville, Md.: U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics, 1997. As of June 21 2005:
<http://www.cdc.gov/nchs/products/pubs/pubd/other/miscpub/nehisrev.htm>
- , “National Employer Health Insurance Survey,” last reviewed January 11, 2007. As of June 21, 2005:
<http://www.cdc.gov/nchs/about/major/nehis/nehis.htm>
- National Research Council, *Analyzing Information on Women-Owned Small Businesses in Federal Contracting*, Washington, D.C.: National Academies Press, 2005.
- NCHS—see National Center for Health Statistics.
- Okolie, Cordelia, “Why Size Class Methodology Matters in Analyses of Net and Gross Job Flows,” *Monthly Labor Review*, Vol. 127, No. 7, July 2004, pp. 3–12. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2004/07/art1full.pdf>
- Oyer, Paul, and Scott Schaefer, “Litigation Costs and Returns to Experience,” *The American Economic Review*, Vol. 92, No. 3, June 2002, pp. 683–705.
- Pearce, James E., “Tenure, Unions, and the Relationship Between Employer Size and Wages,” *Journal of Labor Economics*, Vol. 8, No. 2, April 1990, pp. 251–269.
- Peracchi, Franco, and Finis Welch, “Trends in Labor Force Transitions of Older Men and Women,” *Journal of Labor Economics*, Vol. 12, No. 2, April 1994, pp. 210–242.
- Pivetz, Timothy R., Michael A. Searson, and James R. Spletzer, “Measuring Job and Establishment Flows with BLS Longitudinal Microdata,” *Monthly Labor Review*, Vol. 124, No. 4, April 2001, pp. 13–20. As of March 1, 2007:
<http://stats.bls.gov/pub/mlr/2001/04/art2full.pdf>

- Robb, Alicia, *New Data for Dynamic Analysis: The Longitudinal Establishment and Enterprise Microdata (LEEM) File*, Washington, D.C.: U.S. Bureau of the Census, December 1999. As of March 1, 2007: http://www.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101596
- , “Small Business Financing: Differences Between Young and Old Firms,” *Journal of Entrepreneurial Finance and Business Ventures*, Vol 1, No. 2, December 2002.
- Romley, John A., and Eric Talley, *Uncorporated Professionals*, Santa Monica, Calif.: RAND Corporation, WR-302-ICJ, 2005. As of March 1, 2007: http://www.rand.org/pubs/working_papers/WR302/
- Sorensen, Elaine, “The Crowding Hypothesis and Comparable Worth,” *The Journal of Human Resources*, Vol. 25, No. 1, Winter 1990, pp. 55–89.
- Stinson, Martha Harrison, “Estimating the Relationship Between Employer-Provided Health Insurance, Worker Mobility, and Wages,” paper presented at 10th International Conference on Panel Data, Berlin, July 5–6, 2002. As of March 1, 2007: http://econpapers.repec.org/cpd/2002/109_Stinson.pdf
- Strople, Michael H., “From Supermarkets to Supercenters: Employment Shifts to the One-Stop Shop,” *Monthly Labor Review*, Vol. 129, No. 2, February 2006, pp. 39–46. As of March 1, 2007: <http://stats.bls.gov/pub/mlr/2006/02/art3full.pdf>
- Trejo, Stephen J., “Why Do Mexican Americans Earn Low Wages?” *The Journal of Political Economy*, Vol. 105, No. 6, December 1997, pp. 1235–1268.
- U.S. Bureau of the Census, “Center for Economic Studies,” undated(a) homepage. As of March 1, 2007: <http://www.ces.census.gov/>
- , “Research Program,” undated(b) Web page. As of June 1, 2006: <http://webserver01.ces.census.gov/index.php/ces/1.00/researchprogram>
- , *1997 Economic Census: Survey of Minority- and Women-Owned Business Enterprises*, Washington, D.C.: U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2002a. As of August 25, 2005: <http://www.census.gov/csd/mwb/>
- , *2002 Survey of Business Owners*, Washington, D.C.: U.S. Bureau of the Census, 2002b. As of June 1, 2006: <http://www.census.gov/csd/sbo/index.html>
- , *2002 Survey of Business Owners: Advance Report on Characteristics of Employer Business Owners*, Washington, D.C.: U.S. Bureau of the Census, 2002c. As of June 21, 2005: <http://www.census.gov/econ/census02/sbo/intro.htm>
- , “Nonemployer Definitions,” last revised October 12, 2005. As of June 21, 2005: <http://www.census.gov/epcd/nonemployer/view/define.html>

- , “Business Register,” last revised July 27, 2006a. As of June 20, 2005:
<http://www.census.gov/econ/overview/mu0600.html>
- , “Nonemployer Statistics,” last modified September 19, 2006b. As of June 13, 2006:
<http://www.census.gov/epcd/nonemployer/index.html>
- U.S. Bureau of Labor Statistics, “Employment, Hours, and Earnings from the Current Employment Statistics Survey (National),” undated(a) homepage. As of June 8, 2005:
<http://www.bls.gov/ces/>
- , “National Compensation Survey,” undated(b) homepage. As of March 1, 2007:
<http://www.bls.gov/ncs/>
- , “Quarterly Census of Employment and Wages,” undated(c) homepage. As of June 8, 2005:
<http://www.bls.gov/cew/>
- , *BLS Handbook of Methods*, Washington, D.C.: The Bureau, 1997. As of June 14, 2006:
<http://www.bls.gov/opub/hom/homtoc.htm>
- , “Mission Statement,” last updated October 16, 2001. As of July 5, 2005:
<http://www.bls.gov/bls/blsmisn.htm>
- , “Current Population Survey: Design and Methodology,” March 2002, Technical Paper 63RV.
- , “Establishment Size Data Available from QCEW,” last modified July 20, 2006a. As of May 30, 2006:
<http://www.bls.gov/cew/cewsize.htm>
- , “Survey Methodology,” last modified November 16, 2006b. As of July 5, 2005:
<http://www.bls.gov/ncs/methodology.htm>
- , “National Compensation Survey: Employee Benefits in Private Industry in the United States, March 2006,” August, 2006c, U.S. Department of Labor.
- , “County Employment and Wages,” last modified January 11, 2007a. As of June 15, 2005:
<http://www.bls.gov/news.release/cewqtr.toc.htm>
- , “Researcher Access to Confidential Data Files at the Bureau of Labor Statistics,” last modified January 17, 2007b. As of June 8, 2005:
<http://www.bls.gov/bls/blsresda.htm>
- , “Business Employment Dynamics,” last modified February 14, 2007c. As of June 8, 2005:
<http://www.bls.gov/news.release/cewbd.toc.htm>
- U.S. Department of Health and Human Services, “Medical Expenditure Panel Survey,” undated Web page. As of March 1, 2007:
<http://www.meps.ahrq.gov/mepsweb/>

U.S. Census Bureau (2004), "LEHD Bibliography." Technical Paper, U.S. Census Bureau, March 29, 2004.

U.S. Small Business Administration, "Size Standards," undated Web page. As of March 1, 2007:
<http://www.sba.gov/services/contractingopportunities/sizestandardstopics/index.html>

———, *Mergers and Acquisitions in the United States, 1990–1994*, Washington, D.C.: Office of Advocacy, U.S. Small Business Administration, 1998. As of March 1, 2007:
<http://purl.access.gpo.gov/GPO/LPS20719>

Weiss, Andrew, and Henry J. Landau, "Wages, Hiring Standards, and Firm Size," *Journal of Labor Economics*, Vol. 2, No. 4, October 1984, pp. 477–499.

White, Lawrence J., "Trends in Aggregate Concentration in the United States," *The Journal of Economic Perspectives*, Vol. 16, No. 4, Autumn 2002, pp. 137–160.

Zawacki, Alice M., and Amy K. Taylor, *Contributions to Health Insurance Premiums: When Does the Employer Pay 100 Percent?* Washington, D.C.: Bureau of the Census, 2005. As of June 13, 2006:
http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101737