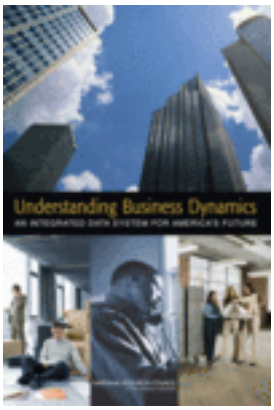

Understanding Business Dynamics: An Integrated Data System for America's Future

Panel on Measuring Business Formation, Dynamics, and Performance, John Haltiwanger, Lisa M. Lynch, and Christopher Mackie, Editors, National Research Council

ISBN: 0-309-66930-8, 200 pages, 6 x 9, (2007)

Understanding Business Dynamics

AN INTEGRATED DATA SYSTEM FOR AMERICA'S FUTURE

Panel on Measuring Business Formation, Dynamics, and Performance

John Haltiwanger, Lisa M. Lynch, and Christopher Mackie, *Editors*

Committee on National Statistics

Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study is supported by a contract between the National Academy of Sciences and the Ewing Marion Kauffman Foundation (Grant # 20040649). The work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (Number SBR-0112521). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

Library of Congress Cataloging-in-Publication Data

Understanding business dynamics : an integrated data system for America's future / Panel on Measuring Business Formation, Dynamics, and Performance ; John Haltiwanger, Lisa M. Lynch, and Christopher Mackie, editors.
p. cm.
Index and bibliographical references.
ISBN 978-0-309-10492-0 (pbk.) — ISBN 978-0-309-66930-6 (pdf) 1. Commercial statistics—United States. 2. United States—Statistical services. 3. United States—Commerce—Statistics. 4. Business enterprises—Statistics. I. Haltiwanger, John C. II. Lynch, Lisa M. III. Mackie, Christopher D. IV. National Research Council (U.S.). Panel on Measuring Business Formation, Dynamics, and Performance.
HF3001.U53 2007
338.0072′7—dc22
2007005756

Additional copies of this report are available from The National Academies Press, 500 Fifth Street, NW, Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

Suggested citation: National Research Council. (2007). *Understanding Business Dynamics: An Integrated Data System for America's Future.* Panel on Measuring Business Formation, Dynamics, and Performance. J. Haltiwanger, L.M. Lynch, and C. Mackie, eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PANEL ON MEASURING BUSINESS FORMATION, DYNAMICS, AND PERFORMANCE

JOHN HALTIWANGER (*Cochair*), Department of Economics, University of Maryland, College Park
LISA M. LYNCH (*Cochair*), Fletcher School of Law and Diplomacy, Tufts University
JOHN M. ABOWD, School of Industrial and Labor Relations, Cornell University
PATRICIA M. ANDERSON, Department of Economics, Dartmouth College
MATTHEW BARNES, Cabinet Office, United Kingdom
STEVEN DAVIS, Department of Economics, University of Chicago
TIMOTHY DUNNE, Federal Reserve Bank of Cleveland
ROBERT M. GROVES, Survey Research Center, University of Michigan, Ann Arbor
SUSAN HANSON, School of Geography, Clark University, Worcester, MA
ROBERT H. McGUCKIN III (until March 2006), The Conference Board, New York
PAUL D. REYNOLDS, Entrepreneurship Research Institute, Florida International University, Miami
MARK J. ROBERTS, Department of Economics, Pennsylvania State University, University Park
NIELS WESTERGARD-NIELSEN, School of Business, Aarhus University, Aarhus, Denmark
KIRK WOLTER, National Opinion Research Center and Department of Statistics, University of Chicago

CHRISTOPHER MACKIE, *Study Director*
THOMAS J. PLEWES, *Senior Program Officer*
MICHAEL SIRI, *Senior Program Assistant*
CARYN KUEBLER, *Research Associate*

Acknowledgments

A long-standing goal of the Committee on National Statistics (CNSTAT) has been to improve the data and statistics that are crucial to accurate and timely economic measurement. In keeping with this history, the Panel on Measuring Business Formation, Dynamics, and Performance is pleased to present its final report. The successful conclusion of this project has resulted from the efforts of many individuals, including but not limited to the panel, whom we wish to thank.

The project was funded primarily by the Ewing Marion Kauffman Foundation. Robert Litan, vice president of Research and Policy, and Robert Strom, director of Research and Policy, initiated the study and provided guidance from the Foundation. Both attended open sessions of meetings to offer their perspectives on the topic and to identify key questions of interest which, in the process, helped the panel sharpen its vision for the study.

Many others generously gave of their time to present at meetings and to answer questions from panel members and staff, thereby helping us to develop a broader and deeper understanding of key issues relevant to the further development of business data systems. The panel especially thanks the statistical agencies; they provide financial support for the project and, even more importantly, allowed the panel access to key personnel with extensive expertise about various data programs. Presenters at the first meeting included Kathleen Utgoff and Jim Spletzer of the Bureau of Labor Statistics (BLS), Frederick Knickerbocker and Ron Jarmin of the Census Bureau, Steven Landefeld and Dennis Fixler of the Bureau of Economic Analysis, and Chad Moutray and Brian Headd of the Small Business Ad-

ministration. Dan Newlon and Cheryl Eavey described the National Science Foundation's (NSF's) interests in the study, focusing much of their discussion on the importance of effective interagency data sharing.

At subsequent meetings, the panel learned a great deal from presentations by Mark Mazur and Nick Greenia of the Internal Revenue Service about that agency's data sharing history, policies, and prospects; Dan Covitz and John Wolken of the Federal Reserve on productivity measurement and the use of financial data on small businesses; Steven Kaplan (University of Chicago) and Josh Lerner (Harvard University) about data sources and research on financing of young and small businesses; Jack Triplett about special data problems for research on the service sectors; and Ron Jarmin, Rick Clayton, and James Spletzer about ongoing business list reconciliation projects at BLS and the Census Bureau. The panel benefited from the comments of Katherine Wallman, U.S. Office of Management and Budget, throughout. The panel also learned a great deal from presentations by Robert Fairlie (University of California, Santa Cruz) about data on the self-employed; Jay Stewart (BLS) on time use data for measuring employment and other business activities; and Martin David (Urban Institute) on data problems for measuring the activity of nonprofit organizations. Maurine Haver (Haver Analytics) and Bruce Phillips (National Federation of Independent Business) expertly presented on the needs of the business community for federally produced data on businesses.

The panel also made an effort to hear about business data developments overseas. At our London meeting, we learned about the development and harmonization, as well as the quality and coverage, of business registers in the United Kingdom from John Perry, Office of National Statistics (ONS). We benefited from a report on the ONS Business Data Linking Project and data access programs from Prabhat Vaze (ONS); a description of user data experiences from Jonathan Haskel (ONS, Centre for Research into Business Activity and Queen Mary, University of London) and Brian Titley (senior economic adviser, director of Performance and Evaluation, Department of Trade and Industry); and commentary about business data systems and research in other European countries from Frederick Delmar, Center for Entrepreneurship and Business Creation, Stockholm School of Economics, and Søren Leth-Sørensen, Statistics Denmark.

The panel could not have conducted its work without an excellent and well-managed staff. In that regard, we appreciate the support of Connie Citro, director of CNSTAT. Senior program officer Daniel Cork, research associate Caryn Kuebler, and senior program assistant Michael Siri provided excellent administrative, editorial, research, and logistical support. The panel also benefited from the work of Christine McShane, Eugenia Grohman, and Kirsten Sampson Snyder, of the Division of Behavioral and

Social Sciences and Education Reports Office, who were responsible for editing the report and overseeing the review process.

The entire panel owes a special debt of gratitude to Christopher Mackie, the panel's study director. During the course of the panel's deliberations, he played an invaluable role in facilitating communication among panel members, identifying studies, reports, and key informants that the panel could draw upon, directing the panel's attention to gaps and inconsistencies in our earlier drafts of the report, and keeping us on schedule. Over the past year, he read and reworked each of the report's chapters multiple times to ensure that the final product was technically accurate yet readable and relevant for a larger audience. All of us on the panel deeply appreciate and have greatly benefited from his knowledge, resourcefulness, organizational skills, and good humor.

The report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Research Council (NRC). The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their participation in the review of this report: Nadin Ahmad, Statistics Directorate, Organisation for Economic Co-operation and Development; Howard E. Aldrich, Sociology Department, University of North Carolina, Chapel Hill; Richard J. Boden, Department of Finance, University of Toledo; Tim Davis, Statistics Directorate, Organisation for Economic Co-operation and Development; William 'Denny' Dennis, Jr., Research Program, National Federation of Independent Business Research Foundation, Washington, DC; Michael Gort, Department of Economics, University of Buffalo; Thomas J. Holmes, Department of Economics, University of Minnesota; V. Joseph Hotz, Department of Policy Studies, University of California, Los Angeles; and Christopher Sims, Department of Economics, Princeton University.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by William Nordhaus, Department of Economics, Yale University, and Harold T. Shapiro, Woodrow Wilson School of Public and International Affairs, Princeton University. Appointed by the NRC, they were responsible for making certain that an independent examination of this report was carried out in

accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring panel and the institution.

Most importantly, we thank the members of the panel for their hard work. This report reflects the collective expertise and commitment of the individual members of the panel. All participated in the panel's many meetings and in drafting material for discussion and, ultimately, for the report itself. Each member brought a critical perspective, and our meetings provided many opportunities for panel members to learn from one another.

Finally the substance of this report and of much work on the topic of business data and statistics in general owes much to Robert McGuckin. Working both on the public- and private-sector sides, Bob contributed prominently to the development of business data. While chief of the Center for Economic Studies at the U.S. Bureau of the Census, he guided development of the Longitudinal Research Database and a broad research program in both statistics and economics. During his tenure, the Center for Economic Studies developed and sponsored research on U.S. business dynamics that has revolutionized the way economists think about and study the U.S. economy. Through his work, economists have learned that the U.S. business sector is incredibly dynamic with a high pace of entry and exit by businesses and an associated pace of job creation and job destruction. The studies he pioneered also showed that much of U.S. productivity growth is associated with this churning of businesses and jobs. He firmly believed that the quality of research based on business data produced by the statistical agencies would improve with greater interaction between outside researchers and businesses and the statistical agencies. As a result, he established the Census/NSF Research Data Center network that enables researchers to access proprietary firm-level data sets for approved research projects that provide new insights into the workings of the economy and the behavior of U.S. businesses. Bob was a member of the panel and participated in early meetings, but died March 12, 2006. We speak for the entire panel in acknowledging his important contributions to this report as well as to the insights from his work that are reflected in this report. We will sincerely miss him as a colleague and a friend.

John Haltiwanger and Lisa M. Lynch, *Cochairs*
Panel on Measuring Business Formation, Dynamics, and Performance

Contents

Executive Summary

The mix and character of businesses in the United States have changed dramatically over the past 30 years. The economy is more integrated and interdependent globally; it has become much less reliant on the manufacturing sector; and new technologies have transformed the nature of work. These transformations have resulted in a highly dynamic economy with outcomes varying over time by sector, region, and segment of society. Business firms, as well as the one or more establishments that comprise them, are constantly changing, with the people who start and run them frequently reinventing their careers. The pace of establishment entry and exit is rapid, especially in the expanding service sectors—additional flux in the economy is created as companies reorganize through mergers, acquisitions, and divestitures. Moreover, it has become more difficult to classify a business as manufacturer, wholesaler, or retailer. Even the physical location of business activity has become more difficult to track, as information technology permits key inputs for some industries to electronically connect into the production process from anywhere in the world. The blurring of boundaries implies that measuring business activity increasingly requires tracking the connection between employers, employees, and independent entities.

The dynamic U.S. economy poses challenges to policy makers at the national, state, and regional levels who seek a more complete understanding of the factors that enhance productivity and innovation, as well as how different sectors and regions participate in the economy. This understanding is provided by the analyses and interpretations—which are in turn

heavily dependent on the federal business data system—conducted by the academic and statistical agency research communities.

Business data collected by the statistical agencies are useful for a broad spectrum of research, policy, and commercial purposes. They provide key building blocks for national and local statistics on income, economic output, employment, productivity, investment, and prices. Beyond these headline statistics, micro-level business data are integral to analyses of job creation and destruction, worker flows, opportunities for economic advancement, firm and establishment entry and exit, technology adoption, innovation, business owner characteristics, outsourcing, interactions among firms, and even the impact of natural disasters on local economic activity. Given these wide-ranging interests, the Panel on Measuring Business Formation, Dynamics, and Performance was asked to develop strategies for improving the accuracy, currency, coverage, and integration of data used in academic and agency research on these topics, as well as data used in the production of key national (as well as regional and local) statistics.

The panel's charge was to (1) catalogue currently available databases, focusing on those produced by the federal statistical system; (2) identify gaps in data sources that impede the production of accurate and timely statistics and that hamper research on business dynamics; and (3) develop recommendations for more effective integration of data sources and for new and improved collection of business data, recognizing legal impediments, survey response rate and burden considerations, and access and confidentiality issues.

Given its historically predominant focus on large and mature businesses, the current federal business data system was designed to provide efficient measures of gross output and net job creation. This is the case because a relatively modest number of well-established businesses account for a large share of the nation's aggregate economic activity. As it stands, however, the U.S. business data system is inadequate for understanding many of the mechanisms leading to greater productivity and innovation or the dynamics of firm and job creation. The drawback to the current approach is that, when business dynamics vary systematically with business size or age, it can yield less accurate, potentially misleading, measures of *changes* in economic activity.

Over the past decade, U.S. statistical agencies have markedly improved the measurement of business activity through the development of longitudinal databases, constructed in large part from administrative records. Private research foundations have also supported improvements in databases on young and small businesses. Nonetheless, substantial data gaps remain.

The panel presents a series of recommendations, all consistent with current norms and standards for use of government data in the United

States, for improving the understanding of U.S. business dynamics. More specifically, we recommend working toward a business statistics system that includes better measurement of young and small businesses, richer analyses of their economic performance and role in the larger economy, and more reliable, timely, and accessible data on entrepreneurial activities. If these recommendations are adopted, this new data system will substantially enhance the capacity of researchers to understand U.S. business and employment dynamics and assist policy makers as they react to major competitive challenges across sectors and geographic regions. While several of the core recommendations could be implemented at low cost (at least in the long run), others would require significant adjustments to current data systems.

Reorientation of the data system as described in this report is needed to address numerous important questions: How, and how much, do young and small businesses contribute to innovation and productivity growth? How important are these businesses in the generation of jobs? Do differences in the entrepreneurial characteristics of business enterprises help explain regional differences in economic performance? And do new and small businesses offer good opportunities for minorities, women, immigrants, and those with less schooling to become part of the economic mainstream? To advance research on the role that new businesses play in the evolution of the economy, richer data are needed on business enterprises, business owners, and the legal, fiscal, and economic characteristics of the environments in which they operate. The "ideal" data system must be capable of integrating data from an array of sources—private and public, business- and household-based, cross-sectional and longitudinal, survey and administrative, national and subnational—that permit business dynamics to be measured in ways that are just now being conceptualized.

BROAD PRINCIPLES

During its deliberations, the panel identified four principles to guide its work and, in turn, the development of its recommendations:

*1. **Confidentiality: The statistical agencies have a responsibility to data providers and data subjects to protect the confidentiality of information that is provided.*** Data collected by the government must be maintained in such a way that identifiable information is not disclosed for administrative, regulatory, or enforcement purposes.

*2. **Public Purpose: Subject to confidentiality requirements, data sharing among government statistical agencies and data access by others should be facilitated when it serves a substantial public purpose.*** Data uses that serve a substantial public purpose include those that lead to improvements

in the quality, breadth, and usefulness of government statistics; provide evidence crucial to informing government policies on social and economic issues; and encourage research that advances scientific knowledge.

3. Targeting Deficiencies: Improvements to data collection should focus first on areas in which policy and research relevance is high but in which statistics needed to inform those policies and research are weakest. This implies building up the statistical infrastructure for measuring business dynamics and collecting information on rapidly growing economic sectors in which the activities of smaller and younger firms are disproportionately important, but for which data coverage is relatively weak.

4. Cost Efficiency: The statistical agencies should give the highest priorities to actions that can be done expeditiously and at low cost. Throughout this report, we identify situations for which more creative use of existing data can be exploited for the purpose of producing useful statistics.

RECOMMENDATIONS

The recommendations in this report are organized into three thematic groupings. The first set of recommendations confronts the need to increase the statistical system's capacity to measure activities of nascent and young businesses—especially those positioned in fast-growing and innovative sectors of the economy—that are central to understanding business dynamics. The second set of recommendations outlines actions to improve the coverage and depth of business data through more effective coordination and integration of existing information sources. These recommendations reflect the need to improve business data while recognizing that statistical agency budgets are tight and that containment of respondent burden is essential. The third set of recommendations is directed toward shifting the legal and organizational environment to accommodate data sharing and confidentiality protections in such a way that enables the kinds of efficiencies envisioned by the panel to occur.

The panel's broadest recommendations for helping guide the plans of statistical agencies are listed here. Steps that can be taken to improve specific surveys and business lists are detailed in Chapter 5.

Expanding Data on Young and Nascent Businesses

In designing a data collection system, nothing is more fundamental than the question of whom to survey. The optimal mix of established business entities to be covered in the statistical system's surveys, censuses, and administrative sources must be determined. For measuring business dynamics, it would be beneficial to reduce the undersampling of those parts

of the business population that are most likely to be in transition and that provide early indicators of the future directions of the economy.

To measure business dynamics more effectively, the Census Bureau and the Bureau of Labor Statistics (BLS) should increase the sampling of younger units in their surveys. This will require that business age be included as one of the stratifying variables and that business lists, on which the surveys are based, cover recent business entrants.

Given the panel's conclusion that essential policy-based research relies on information about new businesses, it follows that key data programs must keep track of how long business entities have existed. Business statistics in the federal system are regularly disaggregated along other dimensions—by firm or establishment size, for example—but very little information is systematically produced or tabulated by age. However, it is clear that a better understanding of dynamic trends in industry evolution, firm entry into markets, and the productivity impact of new firms requires data on business age.

The Census Bureau and BLS should exploit their administrative record systems to produce public-release statistics on economic activity disaggregated by indicators of business age. Readily available business age indicators in these administrative records systems include the application date for an Employer Identification Number, the point at which positive revenues are generated, and the first period with positive payroll.

A focus on publishing statistics by business age would also be compatible with the recent innovations in measuring producer dynamics, such as those developed in association with the Business Employment Dynamics (BED) program produced by BLS and the Statistics of U.S. Businesses (SUSB) produced by the Census Bureau (and the closely related Longitudinal Business Database (LBD) microdata program at the Census Bureau).

Because current data collection focuses on larger business entities and traditional sectors and employment arrangements, activities associated with some of the most interesting and rapidly changing components of the economy are imprecisely, or slow to be, detected. Measurement and analysis of the processes through which businesses are born and grow require going beyond conventional data collection from employer businesses. The most direct way to get at early life-cycle dynamics involves focusing on household or individual units. While there are limitations to household-based data, such as the typical absence of information on business performance, they can be used as a screening vehicle for identifying nascent and young businesses and, subsequently, for generating information on their

transitions to more substantial business entities. Optimally, to form as rich a picture as possible, information on worker and entrepreneur characteristics, self-employment, and household-centered businesses should be integrated in a longitudinal data infrastructure.

Improving Measurement of Business Dynamics Through Efficient Use of Existing Information Sources

Given finite, often tightening, resources, a realistic strategy to improve business data must rely heavily on effective use of existing data collection efforts. Many of the measurement objectives described in this report can be achieved without major new investments; they require only improved coordination of currently available data.

A key aspect of the strategy to make more effective use of existing resources involves overcoming technical and legal hurdles so that administrative data that are routinely collected from (and by) businesses can be broadly exploited. For example, major sources of data on self-employed individuals are administrative—tax return information, such as that contained in Schedule C returns, is particularly important. Use of administrative sources can (1) improve data accuracy, particularly when survey questions require respondent recall; (2) broaden population coverage; and (3) reduce respondent burden by minimizing the amount of information that must be gathered in duplicative surveys. Effective use of administrative data allows surveys to be used in a targeted way when detailed information on special topics is needed.

In order to take advantage of disparate sources, it must be possible to link records at the individual entity level. For measuring business dynamics, it is particularly important to develop a linking strategy that allows construction of comprehensive longitudinal data structures that capture events as they take place over the course of a firm's or an establishment's life cycle. Many of today's surveys and censuses have longitudinally incompatible questionnaires. More weight should be given to the longitudinal uses of these data when survey instruments are created and revised.

The Census Bureau should develop a fully integrated longitudinal household-business data infrastructure from administrative data to serve as a platform for tracking business formation, for integrating household and business survey data for measuring economic activity associated with the business formation process, and for developing samples for new surveys of business dynamics. The integration should include the master household address files, the job frame from linked employer-employee administrative records, and

data for firms (including those with no paid employees, but with receipts) from the Census business register.

An efficient data collection infrastructure also requires that survey programs be well coordinated across statistical agencies. It is not economically efficient to expand the Census Bureau's surveys to include more information, or to collect it at more frequent intervals, when similar data are already collected in BLS surveys. Similarly, it is not efficient to add output and nonlabor input measurements (such as capital investments) to BLS high-frequency surveys. However, periodic measurement of all these concepts on the same questionnaire (and from the same entities) is the only way to identify and correct errors in the estimation of dynamic relationships that occur when the microdata from multiple sources are aggregated for use in statistical products.

BLS and the Census Bureau should jointly develop intermittent topical modules for their business surveys. These topical modules should be designed to allow periodic measurement in the same survey and with the same business sample of variables usually collected in separate surveys and at different frequencies.

Statistical agencies may also be able to improve the accuracy and timeliness of their products by tapping into data systems maintained by businesses. Given that businesses must continually update their own employment, payroll, capital expenditure, and other records, it makes sense to develop conduits from internal reporting systems to government data collections. By recognizing that companies maintain accounting systems associated with day-to-day operations on a high-frequency basis, it may even be possible to mitigate business respondent burden. New technologies (e.g., web-based reporting) will continue to enhance these kinds of opportunities to improve the timeliness and accuracy of collection efforts.

Improving the Business Lists Through Interagency Data Sharing

Four business registers in the United States provide wide-scale coverage of both publicly and privately held businesses: three are maintained by government agencies (the Internal Revenue Service (IRS), BLS, and the Census Bureau), and one is private (Dun & Bradstreet). The registers at the Census Bureau and BLS are the primary lists from which statistics on firm and establishment dynamics are generated. The two main programs on business dynamics—the BED and the SUSB/LBD—are constructed from microdata on establishments in these files. The BLS data come from the Quarterly Census of Employment and Wages program administered by the state Unemployment Insurance programs. The Census Bureau program is

based on IRS filings augmented with data from various censuses and surveys.

The business lists serve a number of critical purposes. They are used to create sampling frames for a wide variety of surveys conducted by the agencies; to benchmark survey data; to publish employment and wage data; and to generate key statistical aggregates, most notably many of the inputs to the national income and product accounts. Comparison projects conducted cooperatively between the Census Bureau and BLS indicate that the business lists do not consistently align—for example, in terms of employment and establishment counts assigned to certain industries, or in their ability to pick up small and new businesses. Inconsistencies in the business lists carry direct implications for the reliability of key business statistics—from gross domestic product, to aggregate employment, to productivity and industrial production—derived at least in part from business list data. In turn, this creates problems for data users. Perhaps most notably, the Federal Reserve's monetary policy is affected when productivity data—calculated using output data from the Census Bureau and input measures (industry employment) from BLS—are inaccurate, because that information factors directly into measured inflation trends.

Continued evolution of the U.S. business data system hinges, to a significant extent, on improving the Census Bureau and BLS business lists. The potential of reconciling the business registers is a highly visible example of what can be gained through effective interagency data sharing. Recent legislation—specifically the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), which allows sharing of confidential business data among BLS, the Bureau of Economic Analysis (BEA), and the Census Bureau for statistical purposes—provides a foundation facilitating the kind of data coordination needed to work toward this goal. However, the Census Bureau is not permitted to share its underlying business list or survey data with BEA or BLS because they are commingled with federal tax information. The panel supports extending CIPSEA to increase the flexibility with which information can be shared among statistical agencies for purposes of constructing a comprehensive business register and for designing special surveys.

Effective coordination of the statistical agency data programs is essential for improving the accuracy, coverage, and timeliness of business data, as well as the efficiency with which they are produced. Before work can progress further to reconcile the business lists, and before data sharing among the three CIPSEA-designated agencies can be fully exploited, the IRS regulations and tax code legislation must be changed.

Measures should be taken immediately to facilitate the expansion of CIPSEA to increase the kinds of information that may be shared

among statistical agencies for the purpose of reconciling the business lists and for the design of special surveys. This expansion of data sharing can be accomplished by: (1) Congress acting to revise Internal Revenue Code Section 6103(j) to extend authorized access of IRS tax information to BEA and BLS; (2) the Treasury Department initiating an update of the IRS regulations, which clarify the purpose and detail specific items that can be shared with authorized agencies; or (3) a combination of both actions.

The goal of the agencies charged with creating and maintaining the source lists should always be to include accurate data on all business units, large and small, new and old. There are, according to the Census Bureau, roughly 18 million nonemployer firms—individual proprietorships, partnerships, or corporations with no paid employees—that account for about three quarters of all firms and a reasonably large fraction (12 percent in 2000) of aggregate U.S. business revenues. Thus, sharing of business data would be quite limited if it did not permit an integration of data on the full range of businesses operating in the economy.

In order to create a comprehensive business list and to generate data that would be useful for studying the dynamics of small and young firms, interagency sharing agreements should extend to data on nonemployers. Data on all sole proprietors and partnerships must also be included, whether they have employees or not.

We believe that a compelling data-driven case has been made, in this report and elsewhere, that reconciliation of the business lists would better serve downstream users—such as BEA in the production of national accounts, the Federal Reserve in carrying out research to inform monetary policy, and the Congressional Budget Office in projecting real gross domestic product growth—to an extent that more than warrants the actions recommended here. The political and legal feasibility of expanding data sharing among the statistical agencies has been enhanced by CIPSEA; indeed, given the uniform set of requirements enacted through CIPSEA, the agencies are now in a better position than ever before to protect data collected for statistical purposes under a pledge of confidentiality.

Increasing the Value of Data Collection by Expanding Use

The statistical agencies rightly view themselves primarily as data producers, and their mission is to do so, maximizing quality subject to budget constraints. While the agencies do maintain skilled in-house staffs, the vast share of research expertise resides elsewhere, at universities or other nongovernmental institutions. It is the intensive use of statistical agency prod-

ucts that creates their high public value and also exposes their strengths and weaknesses, which ultimately feed back to improve the quality of surveys and of the data sets (public use and restricted access) themselves. A full return on the nation's investment in data requires that users have access.

> **The quality of research based on business data produced by the statistical agencies would improve with greater interaction between outside researchers and businesses and the statistical agencies. As recommended in previous Committee on National Statistics reports, statistical agencies, in particular the Census Bureau, should incorporate into their missions a broader interpretation of the criteria for access to data. Specifically, research that informs social and economic policy should be considered a valid reason for accessing confidential data.**

The panel commends recent steps taken by the Census Bureau and IRS to emphasize the importance to public and private decision making of research that takes place at the agency's data centers, and to work out procedures to facilitate streamlined processes for reviewing proposed research projects.

In planning a data system capable of measuring business formation and dynamics, it is essential to keep in mind the needs of users—federal agencies and researchers, as well as businesses themselves. Questions about business activity are frequently made in reference to specific places, ranging from neighborhoods to the entire country. Thus, for many purposes—such as state and local planning—data must be collected and accessible in a way that allows for small-area analyses. Data with precise location identifiers are also needed to document the effects of federal government policies and actions. The impact on local economies of base closures, contract awards, emergency relief, and extending eligibility for unemployment insurance are but a few examples of situations in which the capacity to fully analyze events has been compromised by data limitations. The importance of tracking businesses and people at substate levels has been reinforced by the series of recent natural disasters that have disrupted and redirected many kinds of business and worker activity. Designing useful data systems therefore requires building in the capability to readily aggregate to a range of geographic scales.

The statistical agencies have done a good job of integrating small-area details in many of their data programs. Survey programs should continue to collect, and administrative record systems should maintain, data that enable (1) identification, for authorized purposes, of detailed geographic and sectoral location of business activity, generally at the establishment level; and (2) flexible aggregation of statistics by product, industry, region, county, etc. Because point-level geographic identifiers uniquely identify a site of

business activity, issues of confidentiality arise and access to these kinds of data is typically restricted.

Costs and Priorities

Actions associated with two of the report's core themes involve minimal increases in resources yet have the potential to yield high value. These themes are (1) that the statistical agencies should *maintain* their business registers in a more fully coordinated manner and (2) that statistical agencies should *utilize* the registers to produce new tabulations of economic activity, specifically by business (establishment) age. The recommendations associated with these themes should be given high priority. The report also includes recommendations that would require greater resources or longer term effort to carry out. The rationale behind several of these is that young and small establishments should be given more weight in survey sampling than their receipts or total employment might suggest, because their characteristics change quickly and because they may contribute disproportionately to economic growth. These recommendations are equally if not more important for measuring business dynamics, but the pace at which they can be implemented will be slower. Other recommendations, such as those suggesting more rapid integration of new technologies or more effective use of existing data sources, are offered to encourage long-term efficiency of business data collection. Taken as a set, a major justification for the panel's recommendations is to avoid the costs of "benefits foregone" from the absence of timely, precise data on the mechanisms by which the U.S. economy adapts and grows. In our view, the amount of resources required to provide a more timely, accurate, and complete description of U.S. business dynamics seems like a very good investment of public resources, yielding substantial benefits for future generations.

1

Introduction and Motivation

Maximizing growth, maintaining full employment, minimizing inflation, and advancing the population's living standards are primary economic policy goals. A healthy, market economy is complex, and is characterized by dynamic interactions between businesses and households. The ability to measure these dynamics is of critical importance for understanding the sources of productivity and job growth and, in turn via demand and supply factors, price inflation. Economic policy makers, including the Federal Reserve, rely heavily on timely and accurate statistics on the dynamics of U.S. businesses. Although the U.S. statistical system is a world leader in producing data on business activity, the dynamism of U.S. businesses and the implications of this dynamism for productivity and job growth are only recently becoming evident.

1.1 THE CURRENT SYSTEM

Data on business activity are useful for a broad spectrum of research, policy, and commercial purposes. Business data provide key building blocks for national and local statistics on income, output, employment, productivity, investment, prices, and other economic measures. Data on U.S. businesses are actively used by policy makers at the national, state, and local levels, and by the business community itself in tracking U.S. economic activity. Beyond the public-release statistics that often dominate news reports, micro-level business data are useful for the analysis of productivity growth, job creation and destruction, business entry and exit, the role of

young and small businesses, the characteristics of business owners, outsourcing, interactions among firms, the impact of natural disasters on local economic activity, and other critical issues.

A simple but useful generalization about the U.S. statistical system is that it has been designed to measure levels of business activity—in terms such as business outputs, and business inputs like labor and physical capital—on a timely and accurate basis. Statistical agencies have traditionally focused greater attention on larger, more mature businesses. This approach is capable of producing accurate and cost-effective estimates of aggregate economic activity because a relatively modest number of businesses produce a large share of total output and employ a large share of economic inputs. It is also easier to identify and promptly capture the activity of large, long-established businesses.

But there are drawbacks to this approach; the focus on levels as opposed to growth has led to an underemphasis on young and small businesses. These businesses account for a relatively small share of the level of economic activity but are critically important in measuring and understanding the growth of economic activity. Also when business dynamics vary systematically with business size or age, a focus on larger and more mature units can yield less accurate, potentially misleading, measures of changes in economic activity. For example, given the current focus, the tracking of the response of U.S. businesses to the business cycle is likely mismeasured to the extent that young and small businesses are especially sensitive to the cycle. Equally importantly, a focus on larger and more mature units limits our ability to measure and analyze the early life-cycle dynamics of businesses and to evaluate the factors that impact business formation, selection, and growth. Thus, a full understanding of economic progress and business dynamics requires careful attention to data on younger and smaller businesses.

Over the past decade or so, U.S. statistical agencies have greatly improved the measurement of business activity through the intense development of longitudinal databases constructed from administrative records. For example, the Bureau of Labor Statistics (BLS) has developed the Business Employment Dynamics database, and now regularly produces quarterly statistics on gross job gains and losses by industry, region, and business size class. The Census Bureau has developed the Quarterly Workforce Indicators, which provide measures of worker separations and accessions and related measures of job gains and losses at the local economy level. The Census Bureau has also developed the Longitudinal Business Database (LBD) from its business registers and used the LBD to produce new public-use statistics.[1] The LBD also serves as a micro-level analytical database for

[1]In a closely related program, the Census Bureau has been using its business registers to produce aggregate statistics on the dynamics of U.S. businesses in the Statistics on U.S. Businesses program.

use by researchers at secure sites for valid and approved statistical purposes. Both BLS and the Census Bureau have taken advantage of their new LBDs to help quantify dislocations caused by economic disruptions, such as those caused by Hurricane Katrina and other natural disasters.[2]

Related improvements in databases on young and small businesses have been supported by private research foundations. The Kauffman Foundation, for example, has supported the Panel Study of Entrepreneurial Dynamics, which provides information about nascent entrepreneurs who are considering a business start-up, and the new Kauffman Firm Survey, which will provide information about the early life cycle of business start-ups.

While these developments on the statistical and research fronts represent real progress, the data gaps for young and small businesses are still very substantial. The administrative records data exploited by the U.S. statistical agencies are comprehensive in their coverage, but they provide little depth of information about individual businesses. Administrative records are usually limited to measures of payroll, employment, revenue, industry, business age, legal form of organization, and business location. To quantify and better understand the role of younger and smaller businesses in productivity growth, technology adoption, new product development, and opportunities for economic advance, we need richer types of data on business enterprises, business owners, and the legal, fiscal, and economic characteristics of the environments in which they operate.

1.2 STUDY SCOPE

This study assesses the strengths and weaknesses of existing business databases—their accuracy, coverage, timeliness, richness, and accessibility—and the overall adequacy of the U.S. data infrastructure for measuring and analyzing business outcomes. We compare the current data infrastructure with a feasible ideal system, and advance several recommendations to improve the quality, timeliness, and coverage of U.S. business databases. We emphasize the need for better measurement of younger and smaller businesses, including their evolution over time, and richer analyses of their

[2]Both BLS and the Census Bureau actively used their business registers to study the impact of Hurricane Katrina and to adjust statistics. The detailed location data in their registers permitted measurement and assessment of the businesses impacted by the disaster and to make adjustments to business survey estimates for nonresponse by businesses in the impacted area. The standard approach to nonresponse is often to impute data for the nonrespondent but in the case of Katrina-impacted businesses, BLS, for example, was able to treat nonresponse as nonactive businesses.

economic performance and role in the larger economy; timely and accessible data on entrepreneurial activities is also a major concern. In presenting our recommendations, we prioritize the steps needed to fill the data gaps identified in the report. Given the scarcity of resources for new statistical programs and the desire to keep respondent burden low, key elements of our recommendations stress more effective use of existing data collection instruments, continued development of longitudinal data sets, and greater interagency collaboration in the sharing and linking of data sources within the federal government. We also suggest forms of cooperation between the federal statistical agencies and nongovernmental producers of business data that can yield richer data sets on business characteristics and outcomes while respecting confidentiality requirements.

In considering these issues, we primarily limit our attention to the measurement of business activity in the private, nonfarm sector of the U.S. economy. The agricultural sector involves a number of data sources and measurement issues that are outside the scope of this study. Likewise, the government and nonprofit sectors have distinct features that give rise to different data needs and measurement issues. To keep the scope of this study within reasonable bounds, an assessment of the data infrastructure needs for the agricultural, nonprofit, and government sectors is left for other occasions.[3]

The charge of the Panel on Measuring Business Formation, Dynamics, and Performance is to develop strategies for improving the accuracy, currency, coverage, and integration of data used in academic and agency research on business formation and dynamics, and in the production of key national, regional, and local statistics. The panel's focus is on business formation, young and small businesses, and entrepreneurial activities. Of particular interest are data used to measure and track business entry and exit, job and worker flows, productivity, investment, wages, and prices. Given the keen interest in business formation and growth, the integration of real and financial data that permit the measurement and analysis of financing for young and small businesses is also a key area of interest.

The specific goals of the study are to:

[3]Among the recent work focusing on the nonprofit sector is that carried out by the UN/Johns Hopkins project on nonprofits (http://www.jhu.edu/~gnisp/) and efforts at the Urban Institute's Center on Nonprofits and Philanthropy (http://www.urban.org/center/cnp/index.cfm). Much of the data work on the agricultural sector takes place at the U.S. Department of Agriculture's Economic Research Service. Summaries of leading work on measuring the government sector (for the United States and Great Britain, respectively) can be found in National Research Council (1998) and Atkinson (2005).

- catalogue the currently available cross-sectional and longitudinal databases on business-level outcomes;
- identify gaps in current data sources that impede the production of accurate and timely statistics on business dynamics;
- identify gaps that hamper research on business entry and exit, business evolution over time, interactions among firms, factors that influence business adaptation and growth, the dynamics of the self-employed, useful definitions of business organizations and their scope of operations, job and worker flows, financial and other business-to-business linkages, and the transformation of business activities and organizations; and
- develop recommendations for better use of existing data sources, new and improved collection of business data, and more effective integration of existing data collection projects. The recommendations must respect legal obligations, confidentiality requirements, and access needs, and they must recognize issues related to survey response rates, respondent burdens, and the high cost of longitudinal surveys.

1.3 BUSINESS DATA USES AND CHALLENGES

U.S. statistical agencies collect business-level data to construct national statistics on aggregate income, profits, output, productivity, employment, investment, prices, and other measures of economic activity. There is a high-priority need by the user community to measure both the level and the changes in U.S. business economic activity. While national aggregates have the top priority, statistics on the levels and changes in U.S. business activity are often classified by industry and by location of business activity. Detailed data by industry and location are essential for understanding the rapid pace of restructuring within many industries, the growth of the service sector, the diffusion of advanced technologies and new business practices, and many other aspects of economic development. Measurement of business activity at the regional, state, and county level is also important for a variety of policy questions. Recent natural disasters such as Hurricane Katrina highlight the need for timely information on local business activity.

Challenges in the measurement of business activity at the national, industry, and regional level are many and well known. The ongoing shift in industrial structure imposes its own challenges. There is invariably a lag in the response of the statistical agencies in shifting the focus of measurement to new and growing sectors of the economy. Reflecting their historical importance, much more information is currently collected on the agricultural and manufacturing sectors than the service sector. However, in the last couple of decades the U.S. statistical agencies have responded to this

challenge by developing more comprehensive economic censuses, and annual and quarterly surveys of the nonfarm, service-producing sectors of the economy.

Some large and rapidly growing industries are difficult to measure adequately because of the nature of their activities. It is well known, for example, that the financial services sector poses challenges in the measuring output. Researchers frequently cite measurement of the output of banks which, given the wide range of services offered and the difficulty of quantifying the revenue streams and prices associated with these services, pose serious problems. This challenge has yielded puzzling findings such as negative measured productivity growth in the banking sector despite the presence of apparent and significant technological innovations. A related challenge is the measurement of output and prices of advanced technology products such as computers, semiconductors, software, telecommunications, and new medical treatments. Rapid technological changes and quality improvements can make it difficult for the statistical agencies to adequately capture developments in the market place.

Globalization and the increased role of large, multinational firms is another ongoing measurement challenge for the statistical agencies. Complex, vertically and horizontally integrated firms—with research and development labs and customer service call centers around the globe, and many components produced offshore—make the measurement of business activity increasingly difficult.

Another challenge is reconciling and integrating the measurement of business activity with the economic activity of households. The center point of this reconciliation is employment statistics. Household and business surveys of total employment as well as employment growth rates differ substantially and systematically in booms and recessions. Possible sources of these discrepancies include inconsistent treatment of the self-employed, multiple job holders, and off-the-book workers that might show up in household surveys but not in establishment surveys and administrative records.

Yet another challenge is to generate the core measures of business activity on a timely basis. There is high demand for information on U.S. business activity that is both current and accurate by both the policy and business communities. For example, for the purpose of setting monetary policy, the Board of Governors of the Federal Reserve System requires accurate and timely information up to the latest week and month prior to its Open Market Committee meetings (held every six weeks).

Given all of these challenges, and the potentially conflicting goal of low respondent burden, it might be argued that other needs merit higher priority than better data on entrepreneurial activity, business start-ups, and young businesses. Two basic arguments suggest otherwise. First, it is a

fallacy to argue that since large and mature businesses account for the largest share of activity, they account for most of the changes in aggregate activity. For example, young and small businesses play a disproportionately large role in the creation and destruction of jobs and, in certain circumstances, changes in national and local economic activity. Second, the U.S. economy constantly reinvents itself with new business practices, new products, and new processes. Young and small businesses play a vital role in the ongoing restructuring of the U.S. economy. Failure to measure this role can mean missing much of the story.

The panel put a high weight on the resource constraints for business data collection. Resources for U.S. statistical agencies are limited, and the challenges they face are considerable. Moreover, keeping respondent burden to a reasonable level (and perhaps reducing it) is an important consideration. Given these resource constraints and objectives, we explore ways that existing data collection can be made more efficient through data integration. This means the integration of survey and administrative records data and the combination of data across surveys. Challenges for this approach include legal restrictions on data sharing and access across U.S. statistical agencies. The panel's findings and recommendations have these resource and legal constraints in the background of all of the discussion.

1.4 THE VALUE OF STUDYING BUSINESS DYNAMICS

Longitudinal databases have yielded several insights into business dynamics and the operation of the larger economy. For example, we now know that gross job creation and destruction dwarf net employment changes at the national, regional, and industry levels.[4] Indeed, about one in seven jobs in the U.S. private sector disappears in an average year, and an even larger number of new jobs are created. Much of this job creation and destruction reflects business start-ups and shutdowns. Exits and other deep employment cutbacks at the level of individual businesses translate into job losses for workers and, often, unemployment spells and reduced earnings. Thus, there is a close connection between individual business dynamics and the fortunes of workers and their families.

Better measurement of business activity, especially at young and small businesses, is also essential for addressing other key economic issues. The

[4]Gross job creation is calculated by summing employment gains over new and expanding employers. Likewise, gross job destruction is calculated by summing employment losses over exits and contracting employers.

large-scale turnover of firms, jobs, and workers in the U.S. economy reflects an ongoing process of business responses to idiosyncratic shocks and differential responses to common shocks. Longitudinal studies find that the continuous reallocation of jobs, workers, and capital from less to more productive businesses is an important source of aggregate productivity gains. Younger firms appear to play an especially important role in this process, as suggested by their relatively high and variable growth rates.

Some of the dramatic outcome differences among young businesses reflect high levels of experimentation with different business methods, production processes, organizational structures, new products, and new locations. Preliminary research findings suggest that each cohort of entering businesses is quite heterogeneous, and that entrants often experiment with a variety of new methods, products, and processes.[5] Some of these businesses discover or develop commercially successful innovations, become profitable and expand rapidly, thereby contributing to employment gains and productivity growth. Many other new businesses, however, do not survive the competitive selection process, and they eventually shrink or exit. The available evidence suggests that this market selection process contributes to productivity gains in the sense that less productive, less profitable businesses tend to exit, while the more productive, more profitable firms tend to endure and expand.[6]

These findings about the churning of jobs, workers, and firms and their contribution to productivity gains are preliminary—in part because the measures of productivity available for young and small businesses are so limited. As emphasized above, most of the data on outputs and inputs, including key inputs like capital expenditures on advanced technologies, are collected mainly for large, mature businesses. Hence, the productivity measures available for younger and smaller firms are often quite crude.

As a consequence, many critical questions cannot be adequately addressed with existing data sources. If business start-ups and the postentry dynamics of young businesses play a vital role in economic growth and fluctuations, as suggested by the available evidence, then it is important to measure and understand the factors that influence these dynamics. Likewise, measuring and understanding the factors that influence the decision to become an entrepreneur is important. In addition, relatively little is known about the activities of young and small businesses in terms of busi-

[5]See, e.g., Davis, Haltiwanger, and Schuh (1996); Foster, Haltiwanger, and Krizan (2001); Syverson (2004); and Becker et al. (2006).

[6]See Bartelsman and Doms (2000) and Foster, Haltiwanger, and Krizan (2001, 2002).

ness methods, products, and processes. And little is known about the obstacles to the survival and growth of young businesses. Some researchers argue that access to external financing is critical for the success of business start-ups, but there is little comprehensive research on this issue because of data limitations. The share of privately held firms that obtain venture capital financing is very small, but many of the businesses that have gone public in the last decade or so have had such financing (Kaplan, Sensoy, and Strömberg, 2005). Moreover, young businesses that go public sometimes exhibit especially rapid growth. Because young businesses that offer a promising business model can more readily attract venture capital and go public, the causal connection between financing and growth is difficult to pin down, but the interaction between financial markets and young business dynamics is one key issue that calls for better data.

Some existing data sources do focus on small business financing. For example, the Federal Reserve conducts a survey on small business financing that produces much useful information. The latter survey is a rich data resource for these issues but, with no intended criticism of this very valuable survey instrument, it offers far too little given the above characterization of business dynamics. As described above, business start-ups and young businesses must make decisions about the business model, the product, the process, and the location of activity. These decisions are yielding rich and heterogeneous outcomes on key measures like productivity, investment in physical capital and advanced technologies, and job growth. To understand these outcomes, measures of the latter need to become available (in many cases they are not) and then integrated with information from surveys about financing.

1.5 APPLICATIONS THAT WOULD BE ADVANCED BY FURTHER DEVELOPMENT OF DATA ON YOUNG AND SMALL BUSINESSES

The Federal Reserve is among the most prominent users of business data. Consider the Federal Reserve Open Market Committee (FOMC) meeting at which the current status of the economy is evaluated and its behavior over short and long horizons forecasted. The FOMC uses high-frequency statistics on key indicators like employment, unemployment, and sales to conduct this evaluation and make these forecasts. Start-up and young businesses are volatile and among the most sensitive to business cycle fluctuations. The rapidity with which the economy emerges from recession may very well turn on how well business start-ups and young businesses cope with a changing economic environment. Currently, the employment statistics from businesses are based on the current establishment survey which adds start-ups and young businesses to its sample frame with a lag. More-

over, given that it is a high-frequency survey with substantial nonresponse, BLS primarily uses the employment changes reported by establishments that responded to the survey in both the current and prior month to generate its estimate of employment growth. BLS realizes this shortcoming and has developed sophisticated statistical methods for imputing or forecasting the contribution of business entry and exit. However, they have relatively limited information on which to build such imputations. If real-time data on business entry and exit were available (or with a short time lag) this would significantly improve the ability of the FOMC to detect business cycle turning points.

Next, consider national and local policy planners attempting to evaluate the impact and plan for the future in response to economic dislocation from natural disasters like Hurricane Katrina or from military base closings and realignment. Details on the spatial variation of firms and workers down to the block level are essential to measure the impact of such events. The tracking of business start-ups and shutdowns as well as the growth dynamics of young businesses is vital for evaluating the planning and recovery from the economic dislocation. For example, in New Orleans, tracking business shutdowns that are temporary or permanent, and measuring and tracking the types of businesses that are returning, starting up, or expanding is of critical importance.

In turn, consider the U.S. statistical agencies charged with measuring the activities and productivity of different industries. Anecdotal and their own data collection reveal that there are blurring of boundaries across firms in a variety of ways. Outsourcing of activity implies that some part of the production process is now conducted in a different physical location (either domestically or offshore) and in a different industry. Changes in the employer and employee contractual relationship take the form of increased use of temporary help, personnel service, or employee leasing firms. In both cases, some part of the business start-ups that are observed reflect these changing boundaries of firms. Thus, even for the measurement of the activity of large, mature firms tracking the business start-ups and their activities is of critical importance to measure the activity and productivity of the industry. Failure to capture outsourcing or employee leasing can yield spurious changes in the measured productivity of the firms and the industry as it might appear that the firms and in turn the industry are able to produce the same amount or even more with seemingly fewer inputs.

Finally, consider an academic researcher who is exploring the idea that the financial market deregulations and innovations of the 1980s and 1990s played a fundamental role in the improved U.S. economic performance in the 1990s and in the new century. Testing the hypothesis depends critically on exploring whether financial market innovations permitted greater risk-taking by businesses as even young businesses were able to find financial

backing for risky but high-potential-payoff projects. Both the research community and the policy community have enormous interest in this issue since it impacts regulation and legislation of financial markets in the United States. This issue can only be explored if researchers have access to longitudinal data with information on business start-ups and on tracking measures of real activity like sales, employment, investment in innovative activity, investment in physical capital, the organizational structure of the firm, and the worker mix at the firm. In addition, researchers need information about the sources of financing of these start-ups and young businesses. To make matters even more challenging, the longitudinal business data must also accurately track exits since studies would be biased if the sample included only firms that survived and succeeded. Researchers could only conduct this type of pathbreaking work with substantially improved data on young and small businesses along the lines discussed in this study. Moreover, researchers may only conduct their analyses if they have access to such data.

1.6 THE PANEL'S WORK

During its deliberations, the committee identified a set of principles to guide its work and, in turn, the development of its recommendations. The first principle relates to confidentiality and privacy:

> ***Principle 1:*** *Statistical agencies have the responsibility to data providers and data subjects to protect the confidentiality of information that is provided.*

Data collected by the government must be maintained in such a way that identifiable information is not disclosed for administrative, regulatory, or enforcement purposes. In addition to administrative data, agencies collect information under a pledge of confidentiality for exclusively statistical purposes (National Research Council, 1993, pp. 56-57). Such information may not be disclosed in identifiable form for any nonstatistical purpose without the informed consent of respondents. Statistical purposes include "the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups" (The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)). Nonstatistical purposes include using information for administrative, regulatory, law enforcement, judicial, or other purposes that may affect the rights, privileges, or benefits of a respondent. Avoiding disclosure of confidential data is essential from an ethical perspective, and for maintaining data systems that rely heavily on organizations and individuals to respond to voluntary surveys. Confidentiality protections are intended to minimize respondents' concerns that data will be misused and,

in turn, encourage respondents to provide more accurate information to data collecting agencies and investigators.[7] CIPSEA has gone far in establishing a consistent set of cross-agency guidelines, including penalties for unauthorized disclosure of confidential statistical information.

The second principle relates to the public purpose of statistical agency products, and draws from *Principles and Practices for a Federal Statistical Agency* (National Research Council, 2005a):

> ***Principle 2:*** *Subject to the confidentiality requirements identified above, data sharing among government statistical agencies and data access by others should be facilitated when it serves a substantial public purpose.*

Data uses that serve a substantial public purpose include those that (1) lead to improvements in the quality, breadth, and usefulness of government statistical systems; (2) provide evidence-based analyses of government policies and of social and economic issues; and/or (3) contribute to advances in scientific knowledge.

The rationale for the public purpose principle is straightforward: Government administrative record systems and survey databases generate enormous public value in terms of informing decision makers (including those in the private sector) and are maintained at considerable cost to the public in the form of taxes and the time and monetary outlays associated with complying with reporting requirements.[8] As such, the public is entitled to the full and effective use of these government assets, provided that such uses do not compromise the confidentiality and privacy assurances afforded to respondents.

Data systems should be designed to fulfill, to the greatest extent possible, the needs of users—researchers, policy makers, businesses, and the statistical agencies themselves—conditional on budget and on adhering to pledges of respondent confidentiality. The statistical agencies (BLS, Census Bureau, and the Bureau of Economic Analysis) require data to produce key aggregate income, product, and employment statistics. Currently, economic statistics are generated from multiple sources without an agreed-upon, centrally maintained universe of businesses. As a result, our business statis-

[7] *Private Lives and Public Policies* (National Research Council, 1993) and follow-up reports (National Research Council, 2000, 2005) comprehensively cover data access and confidentiality issues associated with social science data. Chapter 7 of *Private Lives and Public Policies* specifically deals with statistical data on organizations and includes a discussion on sharing business lists.

[8] The arguments underlying the "public purpose principle" are articulated in detail in National Research Council (1993, 2000, 2005b).

tics undersample identifiable business groups in cross-sectional data and produce multiple independent measures of relatively straightforward economic statistics such as the monthly employment and payroll figures. While differences in estimates of economic activity across data sets can generate new insight into the underlying dynamics of our economy, it can also generate unnecessary confusion. Establishing a more harmonized data collection system would reduce this confusion.

In addition to the need for information to monitor aggregate trends, researchers and policy makers also require data to perform microanalyses—on topics such as firm entry and exit; the role of young and small businesses in innovation, economic growth, and job creation; trends in employment and productivity; interactions and linkages among firms/establishments, particularly between large and small ones (e.g., former employees consulting for old firm, small companies selling out to large ones); offshore activities, outsourcing, supply dynamics; and characteristics of businesses/business owners (finances, demographics, nonemployer vs. employer businesses). Throughout this report, we have described how these topics relate to policy. For example, analysts at the Federal Reserve—one of the most prominent sets of users of business statistics—are concerned about measuring output and, in turn, productivity by industry. Some of these kinds of data could be significantly improved simply by better coordination of the business lists residing at BLS and the Census Bureau. Developing analytic capacity within statistical agencies so that data collection and database management systems are also designed for use in "evidence-based policy" may require statistical agencies to reexamine their mission.

Businesses themselves benefit from timely and reliable data as they make employment, production, and investment decisions. When data products help meet business information and planning needs, not only does it contribute to a well-functioning economy, it may enhance the data collection enterprise itself by encouraging higher response rates and participation by businesses. By promoting more effective use of government databases, application of the public purpose principle may produce a more favorable attitude among respondents to government reporting requirements and surveys. Participation in voluntary surveys, as well as compliance with mandatory reporting requirements, may improve when businesses perceive that the resulting databases are put to useful purposes. Principle 3 emphasizes strengthening the weakest data links in the system; that is, prioritizing areas where current federal statistics are least developed:

Principle 3: *Improvements to data collection should focus first on areas*

where policy and research relevance is high but where statistics needed to inform those policies and research are weakest.

In other words, resources should be devoted to data improvement in such a way that the greatest marginal returns can be exploited. For business data, this points to the need for building up the statistical infrastructure for measuring dynamics. In other words we need to collect information on rapidly growing sectors in the economy where activities of smaller and younger firms are disproportionately important, but for which data coverage is weak.

Maintaining data relevance also requires keeping abreast of trends in the quickly changing economic landscape. This requires, for example, improving product codes to keep up with expanding areas of the economy (such as communications equipment and other high-tech components of manufacturing). The Census Bureau has been increasingly developing more extensive product codes for services in response to the dynamic economy in which services have become much more important over time. However, much of the old economy is still embedded in the system. Federal Reserve Governor Randall Kroszner (2006) recently illustrated this point using the example of the product category "broadcast, studio, and related equipment." That product category includes 16 subcategories of product data with modest total shipments (e.g., AM and FM radio transmitters, $103 million; cable-TV subscriber equipment, $41 million; studio transmission links, $18 million). In contrast, the product code "data communications equipment" has no break-downs into subcategories, even though it is a $10.5 billion industry. Kroszner concluded that "the task of updating product lists is resource intensive and time consuming, but it is critical to gaining a more comprehensive understanding of developments in the most vibrant sectors of our economy."

The fourth principle has to do with achieving cost efficiency in data programs:

Principle 4: *The statistical agencies should prioritize actions that can be done expeditiously and at low cost—the low hanging fruit.*

In this report, we identify a number of cases whereby more creative use of existing data can be used for the purpose of producing useful statistics. The idea is to get as much information out of the system as possible for a given level of resource and data protection commitment.

This first chapter has discussed the scope, the objectives, and provided an overview of the case for improvements in U.S. business data with a focus on improved measurement of data on young and small businesses. This chapter has only opened the door to the issues and much remains to be discussed. Chapter 2 takes on the fundamental question of what is meant

by a business. Understanding this question is critical for evaluating measurement of business dynamics. Chapter 3 describes the ideal business data system. The discussion is intended to outline the objectives for further development of the system while at least keeping in the background the many factors constraining the collection of and access to business data. Chapter 4 provides an overview of the current system and discusses the gaps between the (constrained) ideal and the actual system in the United States. This discussion is a natural springboard for Chapter 5 which includes the recommendations of the study.

2

What Is a Business?

What is a business, and when does activity for a business begin? Corporate giants like Wal-Mart, General Motors, and Microsoft are unambiguously ongoing and active businesses by any definition. For other cases, it is less clear: Is a trader on eBay a business? Is a software engineer who is currently employed by a software company, but who is considering starting up a spin-off and has engaged actively in some planning, a business? When does a start-up effort cross the threshold between gestation and becoming an operating business? U.S. statistical agencies typically define a business as an entity that is active in terms of having either positive payroll or positive gross revenue, but many other possibilities exist.[1] As will become clear, even seemingly straightforward criteria involve a variety of issues associated with the measurement and interpretation of business activity and business units.[2]

Realizing that businesses are the major source of jobs in the economy,

[1]For example, the Current Population Survey program states that "a business exists when one or more of the following conditions is met: (1) machinery or equipment of substantial value is used in conducting the business; (2) an office, store, or other place of business is maintained, or (3) the business is advertised by listing in the classified section of the telephone book, displaying a sign, or distributing cards or leaflets or otherwise publicizing that the work or service is offered to the general public" (http://www.bls.census.gov/cps/intmanb1.htm).

[2]For discussion of the measurement issues, useful background reading includes Jarmin and Miranda (2002), Spletzer et al. (2004), and Clayton and Spletzer (2005). For discussion of the conceptual issues, see Dunne, Roberts, and Samuelson (1988) and Davis et al. (2006).

as well as of market goods and services, it is important to be able to identify them in a way that is useful for various research and policy purposes. For some applications—such as in measuring national-level output or employment, industry-specific outputs (like radiators for new cars), or the wide range of goods and services provided by a major retailer (like a single megastore providing food, hardware, electronics, and pharmaceuticals within the same structure)—it may be necessary only to identify activities at the firm level, even if that firm operates establishments in several locations. In many other cases, when analysis is focused on understanding the sources of goods, services, or job creation and loss for a given geographic location, it is important to have data on business activities at the establishment level. Physical locations can be identified at the local or regional level, with attention to growth and change for a given neighborhood, county, metropolitan region, or state, or for the sector in which the unit is classified. Such questions require a focus on business entities that make up a geographically defined economy.

In this chapter, we consider these definitional issues in detail. The first part of the chapter focuses broadly on what a business unit is and what entry or exit of a business represents; the remainder of the chapter asks how such concepts are, or could be, measured in the current U.S. statistical system. In that system, there are a number of different units of analysis from which data are collected (see Box 2-1). These include the establishment, the company (firm), the line of business, and the owner. Some of these concepts are related to legal and administrative entities, while others are related to the productive or geographic structure of the firm.[3] Given these existing data resources and the possibility of incorporating relatively small modifications, we suggest alternative ideas for collecting data on businesses for the purpose of measuring producer dynamics.

2.1 DEFINING BUSINESS UNITS AND IDENTIFYING BIRTHS AND DEATHS

Any procedure for identifying and counting businesses or for providing reliable statistics on business activities is more useful if it begins with a precise conceptualization of a business entity. However, as alluded to above, a number of alternative definitional or conceptual approaches are possible, and they vary in terms of relevance to the production of timely, accurate data on new and young businesses.

[3]Throughout this discussion the focus is on nontax-exempt organizations, although many of the concepts defined could be used to measure business dynamics in the nonprofit sector as well.

BOX 2-1
Key Measurement Units in Business Statistics

Firm—An organization conducting a business under a legal form of ownership. A firm may operate one place of business or more (such as a chain of restaurants) or have no fixed business address at all (such as a firm represented by a self-employed contractors). Firms may take a number of legal forms (and a single firm may operate more than one type of legal entity):

Sole proprietorship—An unincorporated business owned by an individual. Many businesses run by self-employed persons fall into this category. The business may be the only occupation of an individual or the secondary activity of an individual who also works elsewhere.

Partnership—An unincorporated business owned by two or more persons having a shared financial interest in the business.

Corporation—A legally incorporated business as defined by state laws. These organizations may fall into one of several categories (subchapter S, limited liability, etc.).

An important distinction for data collection purposes is that between employer and nonemployer status:

Nonemployer business—One that has no paid employees but (as typically defined) has positive receipts (for the Census Bureau the threshold is $1,000 or more for most industries and $1 or more for the construction industries) and, as such, is subject to federal income taxes.

Employer business—One that has at least one paid employee (importantly, this triggers inclusion in state UI data systems).

Establishment—A single physical location at which (a firm's) business is conducted or services or industrial operations are performed.

Line of Business—A group of establishments under common ownership producing the same basic output (perhaps defined by an industry code). This delineating concept typically lies between the firm and the establishment levels.

Industry—A product grouping that facilitates analysis of the relationship among products—such as those defined by the North American Industry Classification System (NAICS).

Primary source: http://www.census.gov/econ/www/index.html

For taxation and other government purposes, businesses are defined in legal terms. In this context, the focus is on businesses registered as independent legal entities—such as a sole proprietorship, general or limited partnership, or one of several corporate forms (subchapter S, limited liability, etc.)—with the appropriate government agency, usually a state department

of commerce, as a specific type of entity. Any action that would be pursued by the courts related to tax payments, enforcing environmental regulations, antitrust legislation, compliance with Equal Employment Opportunity Commission guidelines, and the like would require dealing with the legal entities. Hence, development and administration of legal requirements and regulations require a focus on these kinds of definitions; it is not uncommon, however, for a single business, coordinating resources and strategy, to create several legal entities to gain tax advantages, compartmentalize risk exposure, or facilitate development of financial support.

Alternatively, business entities may be defined in a more purely economic sense. In this context, any buyer or seller of goods and services (not acting solely as a consumer) conducting activities that influence prices or quantities exchanged in a market would qualify as a meaningful unit of analysis. Such business entities could consist of the full- or part-time effort of a single person or of an organization of thousands acting in concert. There has been substantial effort devoted to identifying the number of businesses (or producers) in specific markets and to measuring the dynamics of producer-buyer exchanges.

Assessment of producer dynamics depends on defining a relevant market, and such definitions can be difficult to refine. This difficulty surfaces, for example, in antitrust cases, where market demarcation is often vigorously contested. However, if one collects detailed information on both the geographic location of production and the products and services produced at each location, then the analyst has flexibility in defining markets and hence in constructing measures of producer dynamics. This also relates to the point that a one-size-fits-all definition of business births and deaths will not provide adequate information over the full range of applications. The opening of a Home Depot in a city may represent a new business birth (of an established firm) in that location while, at the national level, this would properly be viewed as internal corporate growth. The statistical system should provide flexibility in allowing accurate measures of business dynamics at the market level, whether that market is local or national in scope. It would clearly be a major benefit if data could be organized around both legal and productive entities.

2.1.1 Legal and Production-Oriented Concepts

To be useful, the definition of a business should provide clarity for determining when a "productive organization" exists, as well as for indicating precise events associated with changes in status (e.g., firm birth or death transitions). The existence of a legal entity is relatively straightforward: one exists when it has been listed in an appropriate registry. However, there is not a one-to-one correspondence between the existence of a legal entity and

the existence of a productive entity. Many legal entities never become operating businesses; many businesses operate for some time before they become legal entities—particularly those representing self-employment. Likewise, many firms persist as a legal entity long after all productive activities have been terminated. And, as noted above, productive organizations may be represented by multiple legal categories. For these reasons, legal status alone is a problematic and incomplete indicator of the presence of a productive organization.

A focus on the impact of businesses on markets would emphasize entities as they affect the prices or quantities of goods and services exchanged. From this perspective, the presence of economic transactions is a critical criterion for defining the birth or death of a firm. For example, an individual's or team's efforts to initiate a business start-up, but which have not yet involved any economic exchanges, would be considered personal time allocation but not as an action by an economic actor in any market. A strict application of such criteria would indicate the presence of a productive organization only when an external economic exchange takes place, such as an initial payment for supplies, wages paid to an employee, or the receipt of financial support for business purposes. For example, a personal investment of $1,000 from an owner, placed in a firm checking account, might qualify as a market transaction. Such a conception would not require that the "business" receive any income—only that it engaged in economic transactions, such as the purchase of goods or services in anticipation of the sale of a product. This could lead to the categorization of a large number of start-up efforts, in which preliminary small-scale transactions have taken place, as active productive organizations. For some purposes, such as tracking employment relationships, this may be an appropriate definition of a firm birth—even if the salaries and wages were covered by investors and not by revenues generated from the sale of goods or services.

Issues similar to those described above arise when defining a firm "deactivation." Firm deactivation or firm exit might be defined by any number of measures of activity (or rather, inactivity): zero employment, zero payroll, zero expenditures, or zero revenue are all possibilities. A challenge here is that many firms accurately characterized by these criteria may still have substantial assets and continue to function for some time, at least as a legal entity. Indeed, some firms may systematically switch back and forth between active and dormant status; for example, consider a sole proprietor in the construction industry who periodically takes on contracts and, in turn, employs workers on a temporary basis only as needed.

A core theme, then, is that entry and exit are not clear-cut terms, and the appropriate definition may reasonably depend on the issue in question and the measurement objective. If the focus is on job creation, then being payroll active is important. Alternatively, if the focus is on the distribution

and dynamics of sales or revenues, then being revenue active is important. The implication is that some flexibility is required conceptually, as well as in terms of measurement, for defining business existence, entry, and exit.

2.1.2 Administrative Definitions

There are two main administrative systems in the United States that define basic business units—the tax system and the unemployment systems. The former defines the basic legal forms of ownership: the primary tax filing entities for businesses are corporations, partnerships, S-corporations, and sole proprietorships (farm versus nonfarm). Each tax-filing entity is given a unique Employer Identification Number (EIN), and these reporting entities form the basis of the Business Master File (BMF) maintained by the Internal Revenue Service (IRS).[4] This initial assignment of legal entities, as represented by the unique EINs, is fundamental to the collection of data on U.S. businesses. The BMF defines the universe of all nontax-exempt firms in the United States. Not only does the IRS produce statistics through its Statistics of Income program based on the EIN as the unit of analysis, but also the Census Bureau relies on EINs in their data collection programs.

In the U.S. system, a distinction is commonly made between tax-filing entities that are employers and those that are nonemployers. The latter are businesses with positive revenues but with zero employees (e.g., typically the self-employed or sole proprietors). This distinction is particularly important for data collections based on the Unemployment Insurance (UI) system (ES-202 program), as these include only employer firms. The Bureau of Labor Statistics (BLS) oversees the data collected from the state UI systems and organizes a national database of employers based on them. In the UI system, the legal entity is the UI account. However, through the Multiple Worksite Reports (MWR) system, data on each physical location of employment is obtained from UI account holders, and this allows establishment-level data to be constructed.

While the legal concept of a business as defined by the IRS is central to the U.S. statistical system, the EIN as a unit of analysis or as a statistical sampling unit is often inadequate. As noted above, the EIN is an administrative entity that may or may not correspond to a common conception of an ongoing business as a unified, centrally coordinated productive organization. First, the existence of an EIN does not necessarily mean the entity is an ongoing business. For example, an entity with the potential to become a

[4]Many small businesses that are sole proprietorships file under Schedule C and use their Social Security number as their tax identifier. These firms may or may not have EINs. The restriction is that all employers must file for an EIN.

firm may apply for an EIN with the plan of starting a business, but that business may never actually come into being. The EIN would exist but no taxable business activity would be recorded. Similarly, a business that ceases to exist may still retain an EIN but, again, no business activity may be present in the firm. These situations are readily handled by requiring a certain level of revenue, payroll, or employment to classify as an ongoing business. That said, the choice of threshold variable and the unit of time for which the variable is being measured (e.g., monthly employment or annual payroll) can affect whether an entity is viewed as active or inactive. Second, the EINs for a tax entity can change for both economic reasons and for administrative reasons. A company's EIN(s) may change due to acquisition—the company comes under the legal control of another entity—or due to some administrative tax action such as a change in the legal form of the organization. Third, and most important, the EIN-level unit of analysis may not be appropriate for measuring the economic concepts of interest. This is particularly true in the case of complex business organizations whose EIN-based data may span a range of geographic and product markets.[5]

In order to produce detailed statistics of economic activity by geographic location and by industry, company-based data represented by the EIN unit must be disaggregated into finer units of analysis that make both economic and accounting sense. Two units of analysis have typically been used to measure economic activity below the company level—the establishment concept and the line-of-business concept.

The *establishment* is the primary unit of analysis underlying business data collected and maintained by the U.S. statistical agencies.[6] The Census Bureau defines an establishment to be "a single physical location where business is conducted or where services or industrial operations are performed." In the employer universe (which is the basis for many of the surveys produced by the Census Bureau), there is an additional requirement that the establishment must have one or more employees and be in operation during some part of the year. Firms without employees are referred to

[5]In some larger corporations, multiple EINs may exist that relate to the product structure of the firm. In the case of subsidiaries, each subsidiary will have a distinct EIN and report taxable activity for that EIN under the parent company's consolidated report. However, if a similar type of company organizes itself as a division structure instead of a subsidiary structure, then the division would report its taxable activity under the parent EIN (see IRS Publication 1635).

[6]A UI account could also be considered a unit of observation but has some of the same drawbacks as EINs. A UI account can represent multiple establishments but only part of the firm—it is tied to neither an establishment nor a firm definition. Accounts are state-specific, so the operations of firms producing in multiple states can be difficult to link under common ownership. The MWR system, which augments UI account reporting, does enable a measurement of establishment-level activity.

as "nonemployers" and are often treated separately in Census Bureau data collection programs.

The Census Bureau also allows a physical location to be disaggregated into multiple establishments if there are distinct lines of operation producing different goods (typically goods that are in different industries) at the same location. This disaggregation of a physical location is done only for cases in which the activity in at least two production lines falls into different industry sectors and output can be allocated accordingly. This typically involves large vertically integrated production locations. However, this disaggregation of physical locations complicates the picture because a small, integrated physical entity may be treated simply as one establishment while a large integrated one (doing essentially the same things) may be treated as multiple establishments.

BLS employs a similar establishment concept in the Quarterly Census of Employment and Wages (QCEW) program used to support its business register. Although there are some differences in industry coverage across the Census Bureau and BLS programs, the notion of an establishment is quite similar. One definitional difference that does exist between the agencies has to do with the treatment of very small multiunit operations. While the Census Bureau makes no distinction for reporting based on the size of small multiunit establishments, the QCEW allows small multilocation employers (10 or fewer employees in secondary work sites within the state) to file a combined report as a single establishment.

An alternative unit of analysis, used less frequently than the establishment concept, is the *line of business*. The line of business typically lies between the company level and the establishment and is organized around production in a sector or industry. The Federal Trade Commission line-of-business data program (discontinued) required companies to disaggregate their financial data in this way. In some Census Bureau surveys, large firms are asked to report for a particular line of business. A statistical reporting unit based on line of business may represent the activity (and even partial activity) of a number of individual establishments. Alternatively, a firm may be directly surveyed about economic activity for an industry or collection of industries that, again, may span a set of individual establishments owned by the firm. For example, the Annual Capital Expenditures Survey of the Census Bureau surveys firms regarding capital expenditures, but it requires firms to break out the expenditures by industry.

2.2 DEFINING BUSINESS UNITS FOR THE PURPOSE OF MEASURING DYNAMICS

The key challenges for a data system intended to have the capacity to measure producer dynamics involve accurately tracking (1) the entry,

growth, and exit of business units and (2) the employment flows generated by these expansions and contractions and births and deaths of employers (Davis, Haltiwanger, and Schuh, 1996). The business dynamics literature generally has measured job creation and job destruction by tracking changes in employment in individual establishments over time. Data series on job creation and destruction have been constructed by industry, region, and characteristics of producers, such as size and age. One outgrowth of this line of research in the United States has been the development of the Business Employment Dynamics program at BLS (Spletzer et al., 2004) and the Quarterly Workforce Indicators (QWI), a data series published since 2003 by the Census Bureau that offers detailed information on local labor market dynamics. Each of these programs provides timely information using QCEW data on job creation in new and growing establishments and job destruction in exiting and shrinking establishments. The QWI data series from the Census Bureau also provides measures of worker hires and separations classified by worker characteristics, such as gender and age.

A second approach to defining business units takes an industrial organization perspective and focuses on measuring producer participation in a given market or industry. In this approach, participation in a specific product market by a firm is key to defining the business unit. The specific market may be defined along a range of dimensions including, but not necessarily limited to, industry and geography. The market may correspond to an industry or product that is national (or international) in scope (e.g., semiconductors), or it may be much more localized, as would be the case for, say, the services of most restaurants. Under this definition, a firm or an establishment may produce for a single market or for multiple markets, and it is fundamental to be able to measure a firm's participation in each one.

The business unit described in a market-oriented approach may, in some cases, represent a nondiversified firm that operates in a single market. Here, the legal definition of the firm corresponds to our notion of the business unit, and the standard administrative or tax entity is an adequate statistical unit. Examples include a manufacturing firm operating at a single location producing a specialized good, or a service provider (like a dentist) that sells in a local market. In both cases, these are single establishment firms selling in a single market. However, many firms produce in multiple industries and operate across many distinct geographic markets. The link, in these cases, between the business unit and the firm is not one-to-one, and thus the firm definition is inadequate for measuring participation in markets. It is important to recognize that, while such multimarket firms may be relatively small in number compared with single-market firms, they typically represent a substantial fraction of economic activity because of their large sizes.

From a cross-sectional perspective, there are a number of key charac-

teristics needed to identify a business unit operating in a product market. For one, the ownership structure of the assets owned by a firm must be transparent in the data. That is, the ownership links between the parent company and owned assets, such as establishments, subsidiaries, and lines of business, must be clearly identified. This is necessary in order to determine the number of distinct decision makers in the market. For example, a firm may operate two individual establishments that produce for the same market. Under the product market definition proposed above, one would not want to count such establishments as two independent business units if they are under common ownership. Therefore, each business unit must be identified with the parent firm that owns the unit.

In addition to information on ownership structure, it is important to be able to identify the products and services produced by each firm. Traditionally, this need has been addressed by focusing on identifying industries of operation, enumerating products produced, defining classes of customers, or providing information on the selling format. Across sectors of the economy, different information is required in order to classify the economic activity of producers, as well as to identify the market they sell in. The information required from a retailer is different from that required from a manufacturer. For many sectors of the economy, the location of production is central to defining markets and thus business units, as well. Retail, service, construction, and even manufacturing firms often sell in local markets. For firms in national markets, knowing where they produce may not be critical to defining the business unit; however, it may be important for understanding the impact of the business unit on local input markets (e.g., labor markets). Finally, along with identifying the activities that a producer undertakes, a measure of the economic importance of the activity is required. Shipments, revenues, or sales volumes are likely candidates, although such information is less readily available at the establishment level than it is for employment and payroll.

2.2.1 Producer Dynamics

The ability to measure the entry and exit of market participants, along with changes in the performance of incumbent producers, is fundamental to understanding how markets evolve. Again, our definitions are centered on participation in a product market. An entrant is defined generally as a firm that is producing in a market in the current period (quarter or year) that was not producing in the market in the prior period. An exit is similarly defined as a firm that was producing in a market in the prior period but that is not producing in the market in the current period. For many purposes, one would ideally like as high-frequency data on market entry and exit as possible, perhaps quarterly or, at a minimum, annually. However, it must

also be recognized that attempts to generate data at a higher frequency for entering firms may come with a cost. Identifying the industry, geographic location, ownership, and other detailed characteristics of a business often requires surveys and therefore takes time; at the time of entry, less detailed or accurate information may be all that is available.

On the entry side, it is important to consider both de novo and existing firms. A de novo entrant is a *new firm* producing in a market and is what comes to mind when thinking about entrepreneurial processes. In such cases, the firm did not produce in any other markets prior to its entry into the market of interest. Many single-location births or new nonemployer firms fit into this category. However, another important source of entry into markets is existing firms diversifying into new markets (Dunne, Roberts, and Samuelson, 1988). These existing firms may enter a market by shifting existing assets from one industry to another or by creating a new production facility to service the market. In these cases, the firm may be established, but it is a new participant in a specific market or sector. What is fundamental here is that, in order to fully measure entry, one must be able to identify not only de novo entrants (new firms created for this market), but also the expansion of existing firms participating in a market new to that firm. Similarly, the exit of a producer from a market may correspond to the death of a firm, or it may simply represent a shift in the mix of production occurring at an ongoing firm. Under this definition, the closure of a plant or store may or may not represent exit from a market. These differences in the types of exit also should be distinguishable in data used to study producer dynamics.

In order to measure de novo entrants, one needs reliable and timely information on new firm formation. This information typically comes from administrative sources, such as those underpinning the tax and UI systems. Identifying when existing producers diversify into new markets is in one way more complicated, as it requires longitudinal information on the distribution of economic activity (e.g., along industry or geographic dimensions) of incumbent producers. These definitions of entry and exit require that the ownership of business units is tracked accurately, that industry and product coding is uniform, that the output is measured consistently and, sometimes, that the location of business units is tracked.

Regarding the ownership issues, each business unit must be identified with the firm that owns it at any given point in time, and changes in ownership must be accurately tracked. Ownership and administrative changes that do not affect the business operation but do change the legal entity must not be mistakenly picked up as an exit of an existing firm and the entry of a new firm. Vale (2006) found that an important difference affecting the international comparability of business start-up statistics internationally relates to how countries treat administrative and ownership

changes in the data. Reactivations and reregistrations of businesses can result in substantial overestimation of business start-ups. Baldwin, Beckstead, and Girard (2002) report that business start-up rates are substantially overestimated, both in numbers and employment shares, when one relies simply on tracking firms in a business register. Many new start-ups are, in fact, unlinked acquisitions. However, ownership changes that affect the number of competitors in a market need to be reflected in the data. A horizontal merger between two competitors results in a reduction in the number of independent business units in the market and thus represents a form of exit (but not deactivation of a production unit).

Detailed information on business location may also be vital, depending on the industry or issue under study. Clearly, in such industries as retail and construction—and sometimes even in manufacturing—products markets are local in nature. In order to characterize the entry and exit of firms into these typically narrow geographic markets, detailed information on the location of operations is required. Alternatively, for some industries, such as consulting and perhaps information technology services, the location of the business unit may be effectively undefined. On balance, however, it is probably useful to think of business location as an important characteristic for helping to link business units to specific markets. Many applications require that longitudinal data on the locations of a firm's business activities correspond to the existing definition of an establishment used in current business registers.

Given the market-oriented emphasis in our conceptual definitions of business units, it is clear that high-quality, consistent industry and product coding are also required. While measures of entry and exit of producers in a market should not be driven by changes in industry coding systems, actual shifts by a firm out of some industries and into others should be reflected in data on the creation and destruction of business units (e.g., lines of businesses) within the firm. This requires, among other things, improving and updating product codes to reflect an increasing share of economic activity in dynamic service areas and high-tech manufacturing. This kind of accuracy can be difficult to achieve for new and very small firms because they are constantly changing; thus, one cannot expect to immediately have fully up-to-date coding for some fraction of the business population.

With respect to measuring the dynamics of ongoing business units, precise, standardized records of the same businesses over time are clearly a major asset. As discussed above, inability to accurately track incumbent businesses invariably leads to measurement errors in business birth and death statistics. For these ongoing entities, some measure of size in the market—such as sales, revenue, or output volume—is required so that the growth of incumbent firms can be tracked.

Beyond the period-to-period transitions of firms, analysts of producer

dynamics are often concerned with tracking the history of producers. For example, it is well established that the growth process for young producers differs from that of older producers. However, this process is quite different for new business units that are owned by older firms compared with de novo firms created to enter the market. New businesses owned by older firms are typically larger and have less volatile growth rates and lower failure rates than de novo entrants. In order to develop rich histories and summary statistics on firm dynamics, information on both the age of a business unit and the age of a firm that owns the business unit is useful. Age should typically be defined as an establishment age, and this should be tracked through changes in ownership. Ownership or administrative changes should not generate a new establishment age.

For some purposes, firm age is important but more difficult to measure. Consider, for example, the purchase of an old establishment by a younger firm, in which case the latter becomes the owner if it is the ongoing legal entity. Alternatively, when an older firm purchases a young establishment, the latter should be assigned the age of the purchasing firm age after the transaction. This is similar to a new establishment being opened by an old firm. The establishment age is set to when it opens, but the firm age is based on when the firm opened its first establishment. The key point is that establishment age is clear-cut, defined by the initial period in which it appears in the sample frame. Firm age requires a more detailed enumeration of cases, and rules can be established for this purpose.

2.2.2 Identifying Nascent Businesses

For the purpose of producing many economic statistics, the focus of business data collection must necessarily be on fully operating entities. That said, business creation processes are of great interest, and the conceptual framework for constructing data on these entities is far less developed. In thinking about firm creation, it is useful to begin by defining two key transitions: that which occurs when one or more persons begin to mobilize time and resources to implement a new firm—the beginning of the gestation or firm creation process—and that which occurs when the start-up effort can be considered a going business. There are, at present, no ongoing federal data collection programs that focus on tracking either transition in a representative sample of household or business populations. The Current Population Survey has been used to identify individuals reporting a sudden increase in time devoted to self-employment activities, an indicator of individual business creation activity. But the operational definition—identifying an increase in effort from two consecutive monthly reports of 15 hours per week on new self-employment initiatives—has an unknown relationship to the actual provision of goods and services, job creation, or the

implementation of a going business (Fairlie, 2006). It may be that modest adjustments to the Current Population Survey or other ongoing household-based surveys, such as the American Community Survey program, would provide a basis for estimating the participation of U.S. adults in activities associated with the creation of new businesses.[7]

An ongoing research program—the Global Entrepreneurship Monitor (detailed in Appendix A), which collects data on various aspects of entrepreneurship through a series of coordinated household surveys in a number of countries—has produced operational definitions of start-up transitions. Representative samples of adults have been used to locate individuals that appear to have initiated the creation of a new firm, either on their own or for their employer. Three criteria are used to identify nascent entrepreneurs: (1) they consider themselves involved in new firm creation, (2) they report being actively engaged in behavior to create a new firm, and (3) they expect to own part of the new firm.[8] To separate those in the start-up process from those owning and managing new firms, two criteria have been developed to represent the "firm birth" transition. The operationally more complex one is to identify those that report positive monthly cash flow covering all expenses and salaries for more than three months; a slightly simpler version identifies those reporting salaries and wages paid to the owners for more than three months. There is some correspondence between these criteria and those commonly used to identify new entrants in business registers maintained by the IRS, BLS, Census Bureau, and Dun & Bradstreet. However, inclusion in these registers takes place at different stages of the firm creation process, reflecting considerable variation in the sequence of major start-up events—for example, significant financial investments, hiring the first employees, expenditure on capital and operations, initial sales, and initial profits (Reynolds, 2007).

It has been suggested that, for venture capital-sponsored firms, revenue generation happens, on average, later in a firm's start-up phase than employment creation (Kaplan, Sensoy, and Strömborg, 2005). In such cases,

[7]Substantial work on effective, cost-efficient procedures for identifying nascent entrepreneurs in household surveys has been completed as part of the University of Michigan administered the Panel Study of Entrepreneurial Dynamics (PSED) research program (http://www.psed.isr.umich.edu). Screening completed to identify the PSED II cohort in the fall of 2005 involved adding a 2-minute module to 34,000 household interviews to locate 1,200 active nascent entrepreneurs (a yield of about 35 nascent entrepreneurs per 1,000 households, and about 2,000 minutes of interview time).

[8]These criteria have been employed in dozens of samples in the United States and in over 40 different countries in all stages of development. For U.S. applications, see the appendices in Gartner, Carter, and Reynolds (2004); for cross-national procedures, refer to Reynolds et al. (2005).

employment-based data on existence would measure entry earlier than would the product market participation approach discussed above. Others have emphasized the fact that revenue generation by small firms often precedes the hiring of the first employee and, in fact, many firms never hire an employee through their entire lifetimes (Davis et al., 2006).

While the product market approach leads to a focus on firms from the birth event forward, it has become clear that there is a lack of information and attention to the mechanisms and processes that precede it. Indeed, it is now clear that a substantial amount of time ("sweat equity") as well as resources (mainly informal funding) is absorbed by start-up efforts that never become operating businesses by any criteria. In addition to substantial intellectual interest, a more complete understanding of the start-up process itself has substantial implications for regional economic policies as well as personal career planning.

2.3 CONCEPT VERSUS EXISTING DATA COLLECTION

In the current statistical system, the unit of analysis that maps most flexibly into measurement needs is based on establishment reporting. Under ideal conditions, information on the range of products or services produced at each business location is also provided so that economic activity can be classified into the relevant market(s). The data collected in the quinquennial economic censuses by the Census Bureau have this product- and business-type detail for surveyed establishments.[9] In addition, the data collection procedures of both BLS and the Census Bureau allow establishments to be disaggregated into distinct reporting units if they are important producers of multiple products.

A weak point embedded in both agency procedures has to do with the collection of data from small establishments owned by multiunit firms. The annual Company Organization Survey (described in Appendix A) omits small multiunit firms; thus the Census Bureau captures only the entry and exit of such smaller multiunit establishments owned by small multiunit firms in the quinquennial economic censuses. The MWR do not require an employer to disaggregate the data by establishment until 10 or more employees work away from the parent location. In each case, when dealing with small multiestablishment firms, it may be difficult to measure producer dynamics at the establishment level, especially at high frequencies.

[9]Even in the economic census, a significant number of smaller establishments are not sent survey forms and their data come from administrative record sources. These administrative record sources do not provide disaggregations of output by detailed product or service provided. Output of such establishments is assigned entirely to the industry classification of the establishment.

A key requirement of the ideal business unit definition proposed above is that a measure of output be collected by each business in each period. In practice, this would be expensive, and it is not clear that it would be practical. Certainly, greater use of IRS data on sales and less reliance on administrative data in the economic census would be helpful on this score. Ideals aside, it still must be pointed out that, within the current U.S. statistical and administrative data systems, output data are collected in a very heterogeneous fashion across sectors, across time, and across different types of producers. Every five years, through the economic census programs of the Census Bureau, data on revenue and sales are collected for a wide range of sectors at the establishment level. At an annual frequency, data on revenues are collected by tax authorities on an EIN basis, but these data are not available to all the statistical agencies.[10] Moreover, for larger organizations, revenues at the EIN level are not particularly useful for measuring producer dynamics at the market level.

To be sure, the Census Bureau has a large number of surveys that collect information on output, revenue, and sales that are more timely than the economic census data. Surveys in manufacturing, services, and retail cover large segments of the economy and provide data at annual and subannual frequencies (quarterly and monthly). However, this coverage is less comprehensive, particularly at the higher frequencies, and the surveys focus on providing accurate estimates of sales or output at a point in time at the aggregate level. The design of these surveys focuses on larger units and is not geared to produce dynamics statistics, which require high-quality information on smaller firms and accurate tracking of individual business units over time.[11]

The administrative data generated through the UI system measure employment and payroll systematically for a large number of sectors, at a high frequency (quarterly), and at the establishment level, but they do not contain information on output. These data facilitate production of high-frequency information on producer dynamics at the establishment level

[10]The Census Bureau has for many years received annual extracts of business and demographic data from the IRS. The business data cover all types of tax returns, including those of sole proprietors and tax-exempt organizations, and include a number of items, including sales, total income, and other financial variables such as total payroll and number of employees from the employment tax returns. In economic census years, more tax items are requested (e.g., the description of business activity for sole proprietors), but some of these require special processing by the IRS (not otherwise captured in the master file system) and entail significant costs, which are passed on to the Census Bureau. The annual programs tend to provide the Census Bureau with items already captured by the IRS, which keeps costs down.

[11]A good discussion of the challenges involved in producing statistics on producer dynamics from survey-based data is found in Davis, Haltiwanger, and Schuh (1996).

using employment or payroll data. In these data, establishments are classified by industry, but there is no information available on product detail. Good geographic detail is provided down at least to the county level (and is increasingly available by latitude and longitude). However, defining market participation with these data is somewhat more difficult, since ownership links across establishments are not comprehensive. Under the definition of market entry proposed above, the opening up of a new operation by an existing producer may or may not constitute entry, depending on whether the existing producer is already an incumbent in the market. Because ownership links are not comprehensive in the UI data (especially across states), it may not be possible to classify entry and exit by existing firms in the manner proposed above for certain industries. However, these data do provide the ability to generate establishment-level statistics on entry, exit, and incumbent growth.

A promising new data source that allows for the measurement of producer dynamics is being developed by the Longitudinal Employer-Household Dynamics (LEHD) program at the Census Bureau. These data have some distinct advantages over the existing data sources with respect to measuring and tracking business unit dynamics. LEHD researchers have linked employer data with employee data from a set of states in the UI system. The capability to track both employers and employees allows for more detailed data to be constructed on job creation and destruction. It also allows for the development of measurement algorithms to improve identification of the entry and exit of business units and to relate these new or closing business units to continuing business units by tracking the flow of workers across them.[12] These data have the potential to better describe the entry and exit process of multiindustry and multilocation firms.[13] Moreover, by linking these data to the Census Bureau business register, comprehensive ownership links across establishments can often be constructed.[14] This will generally allow data users to distinguish between de novo and experienced firm entry and exit.

[12]See Benedetto et al. (forthcoming) on this topic.

[13]Employer-level data are at the establishment level, whereas the employee-level data are at the UI account number level. The LEHD program has developed a multiple imputation process to allocate employees of UI account businesses that have more than one establishment to the establishment level, so that worker and job flows at the establishment level can be measured.

[14]Not all of the multifaceted data integration that takes place within the LEHD is based on stable entity identifiers. Employer reports from the QCEW (ES-202) are filed for each establishment; wage records generally originate with UI accounts, and only some states add the reporting (establishment) in the case of multiestablishment firms.

This discussion of the measurement of business activity and entry and exit has focused on the measurement of core concepts like revenue, payroll, employment, product mix, and location. There are other measures of business activity that are important in their own right, as well as for quantifying business entry and exit. One example is physical capital and expenditures. A new business entity may have capital expenditures and assets prior to the onset of revenue, payroll, and employment. As such, the measures of capital expenditures or assets could be used, by some definition, to identify business entry. In addition, measurement of the means by which business activities (for capital expenditures or other expenditures) are financed is extremely important for understanding the processes underlying business formation and development.

2.4 CONCLUSIONS

In this chapter, we discuss a number of different definitions of businesses and how the definition must be appropriate for the use to which it is being applied. If a business frame is maintained with detailed information on the geographic location of production, the products and services produced at each location, the ownership structure of the firm, and the ability to track production units over time, then the measures of producer dynamics outlined above can be constructed. The unit of observation that comes closest to providing the necessary features—and the one that often is and should be used in most statistical agency programs—is the establishment definition. Establishments are physical locations in which economic activity is assigned industry code(s) and whose ownership relationships can be identified. Establishment data with features such as those described above provide the overall flexibility required to measure producer dynamics across a wide range of firm types, from small start-up firms to large expanding and contracting ones. The opening of a new firm and the opening of a new branch of an existing firm can also be distinguished with such data.

The establishment definition also allows flexibility in measuring producer dynamics along geographic dimensions. For some applications, local data on entry, exit, and growth are required. Establishment-level data allow for the construction of such statistics, whereas company or line-of-business data generally do not. Detailed industry coding of establishments would enhance the types of statistics that can be produced. Ideally, it would be useful, for certain industries and types of producers (large firms), to have information on the range of industries (products) in which an establishment operates. Finally, there is justification for using input measures (i.e., employment or payroll) or output measures (i.e., revenues or shipments) as the basic measure to judge economic activity and to create measurement thresholds for defining an active business. An output-based litmus, such as rev-

enues, has an advantage in that it is a basic measure of both employer and nonemployer business activity and is a likely choice to measure aspects of producer dynamics for all segments of the business population.

The bottom line is that establishment data are integral to a flexible business data system, and the statistical agencies should resist giving this up simply because survey respondents do not necessarily organize their data in this particular format. Furthermore, calling mainly for use of the establishment concept in defining a business is a practical statement that aligns with what the statistical agencies can reasonably collect: data on location, industries of operation, ownership links, and economic activity. The framework is already used to generate statistics on producer dynamics; its full development requires that what is already measured (surveyed) in economic censuses should be extended to a larger set of firms in the business population on an occasional basis.

3

The Ideal Business Data System

3.1 GUIDING DESIGN PRINCIPLES

Business data serve many purposes and are relied on by many users. To meet the wide range of needs, a business data system must be flexible along a number of dimensions.[1] For measurement of economic activity, it is important that it be designed with the capacity to disaggregate data at different units of analysis (for example, establishment versus business line versus firm level) in multiunit firms. As noted in Chapter 2, the appropriate unit of measurement varies by sector and type of activity being examined. Furthermore, in order to be able to track employment and other trends geographically, it is critical that data can be disaggregated by location. To capture sectoral trends, the data system must be designed with the capability to follow individual businesses over time, even in cases where changes in the nature of a firm's business or in capital ownership occur. Finally, an ideal data system would be constructed to facilitate the development of modules that allow linkages between business registers, existing surveys, and administrative data. As we describe below, allowing for linkages across sources can help minimize respondent burden as well, and increase the usefulness of final data products.

[1]This theme—that it is important for a business data system to be flexible and to permit "drilling down" from aggregate statistics to the sectoral and firm level—has been explored in the literature. See, for example, McGuckin (1992, 1995) and Becker et al. (2006).

3.1.1 Recognizing and Responding to Multiple User Needs

Among the users of data on business formation and dynamics are federal agencies, researchers, and businesses themselves. One way to increase the probability that final data products will meet demands is to involve users in survey design and data collection strategies from the start. Ideally, statistical agencies should facilitate front-end collaboration with the academic and business communities and not rely solely on postsurvey follow-up. Considering user needs up front should increase the relevance of final data products.

Of course, users' data needs vary enormously. Even among the statistical agencies, whose missions are well documented, this is the case. For example, construction of local-level employment statistics relies more heavily on data for young and small firms than does production of the national economic accounts. Certain organizations—perhaps most notably the Small Business Administration, charged to "aid, counsel, assist and protect the interests of small business concerns"—are focused almost entirely on such data. Another example is the Census Bureau's Economic Planning and Coordination Division (EPCD), which is charged with, among other things, editing and publishing the Census Bureau's nonemployer statistics. They have a number of information needs that could be addressed by a regular survey of nonemployer entities. One example of an information gap that has been problematic for EPCD over the years is the extent to which employee leasing or other nontraditional employment arrangements are utilized by nonemployer entities with very large receipts.

Beyond the agencies, mayors, other local government leaders, and chambers of commerce need information for their cities and surrounding metropolitan areas, while governors need information for their states and counties. Urban and regional planners and congressional representatives require business data aggregated to areas of different sizes. Local business owners take advantage of information about their competition at the local, even neighborhood, level for purposes of planning various aspects of their operations. Ideally, data should be collected and made accessible in a way that helps business owners answer a broad array of operational questions: Where are my customers and potential customers located? Where is the competition located? Where do my employees and potential employees live? Where should I locate my stores, offices, and plants? How much should I produce? How much should I order? How much should I hold in inventory? How should I set my prices? What is the best way to promote my products and services? How much should I pay my employees?

Researchers, policy makers, and businesses use information, disaggregated to various levels of granularity, to plan and set goals for economic growth, to track progress against goals, to identify pockets of underper-

formance, and to devise new programs and corrective actions to improve performance. For many of these purposes, the ideal federal statistical system would facilitate the production of data for small domains—that is, localized geographic areas, specific industries, or any other defined set of firms for which survey data might be desired, even though the corresponding sample sizes may be too small to support direct estimation.

Recent events—such as the aftermath of September 11, or natural disasters such as Hurricane Katrina—have highlighted the need for information on business activity at regional, state, county, and very local levels. Ideally, one could measure not only the impact on economic activity of the initial destruction, but also follow business dynamics during recovery periods in a way that allows the impact of relief efforts to be assessed. Similarly, local area data are an ideal input to many decision processes, such as federal defense office realignments, base closures, and even military reserve unit call-ups, as well as to the analysis of the resulting economic impact of such decisions.[2] Designing data systems with user needs in mind therefore means building in the capability to readily aggregate to a variety of geographic scales, some of which are unknown a priori.

For maximum flexibility in aggregating information, each business establishment should be assigned a unique identifier, where applicable, that records the location as a specific latitude-longitude coordinate (a point location); having the street address of a business location is sufficient, as that information can be linked to geographic coordinates. Because the street addresses of establishments and firms are known and readily accessible to owners and managers, providing this information should not be burdensome to most businesses. For certain research and policy questions, noting whether a business operates from a home location or not would also be useful.

Because a point-level geocode such as a street address uniquely identifies a site of business activity, issues of confidentiality arise with the release of such data. Detailed information on business locations is typically public but, even so, inclusion of unique geocodes in combination with other variables in a database can compromise confidentiality. Yet, because there is such demand for data that enables aggregation to any one of many possible geographic units, a tension is created between user needs and respondent confidentiality; this issue is discussed in Chapter 4.

[2]For example, research on the effect of military reserve deployments and activations on employment at the county level (e.g., Loughran, Klerman, and Savych, 2006), reveals that data limitations make it difficult to assess the impact of their actions on smaller firms and the self-employed.

3.1.2 Managing Respondent Burden

A critical element of an ideal business data system is the capability of managing the burden to businesses of responding to federal surveys. Firms have an obligation to participate in surveys; their participation is a prerequisite to the production of high-quality data that are, in turn, essential for the analyses and planning required for an efficient, productive economy. However, one reason given for nonresponse in surveys by employers is that they take too much time, especially for smaller businesses. So, to ensure accurate representation of economic activity, data collection should be designed in a manner that minimizes the burden on information providers.[3] One aspect of burden management is to avoid asking for unnecessary or redundant data. This highlights the need to make as much use of administrative data as possible. It is important to consider whether there are currently underutilized sources of administrative data that could contribute to the goal of measuring business formation and dynamics while minimizing the survey response burden.

Nongovernment data sources may also play a role in an ideal data system. Given the prominence in the U.S. economy of payroll processing firms, it may be possible to obtain data from a large number of businesses while directly approaching only a few. The firm ADP, for example, claims that it handles payments for 1 in 6 private-sector workers (http://www.adp.com). Adding in the next few biggest processing firms would expand coverage considerably further. In addition, the use of such processing services is not limited to large firms, making it a potentially lower cost way to collect information on smaller ones. ADP offers services for businesses with 50 or fewer employees, major accounts services for firms with 50-999 employees, and national account services for firms with 1,000 or more employees. Potentially, if confidentiality concerns can be addressed, a company such as ADP could provide employment numbers per payroll period in a consistent format for each of its client firms. Given that the smallest firms are the most likely to realize proportionately large benefits from

[3]Although there is a substantial amount of research on the views of individuals on confidentiality, data sharing, and respondent burden, much of it conducted for the decennial censuses. Information about business views on these issues is much more limited. Qualitative research by Willimack and Nichols (2001) suggests that "large companies generally supported data sharing among statistical agencies under well-specified conditions and with rigorous security and confidentiality provisions in place. They only saw value in this sharing, however, if it reduced the reporting burden placed on them" (p. 1). The authors suggest that redundant external data requests, especially for larger firms, appear to make the prospect of data sharing by designated statistical agencies preferable.

outsourcing payroll and human resource functions, payroll processing firms may be able to provide a timely picture of small business dynamics. At a minimum, data collection systems should be structured in a manner similar to those used by these payroll processing firms in order to simplify the survey response by businesses. Whenever possible—that is, when businesses keep records in the same way and with the information that government wants—responding to government data collection requests should be made no more difficult than gathering the information necessary for carrying out basic administrative functions for the firm. This is more likely to be the case for information such as payroll, but less so for information relating to, say, energy consumption or investment.

One way to facilitate the collection of administrative data is to encourage increased adoption of extensible business reporting language (XBRL). XBRL is a member of the family of languages based on XML, or extensible markup language, which is a standard for the electronic exchange of data between businesses (see www.xbrl.org). In terms of U.S. government data collection, XBRL has already made some inroads with regulators. For example, the Federal Financial Institutions Examination Council uses XBRL for quarterly bank reporting. In terms of broader types of data collection, tax authorities in several countries (e.g., the United Kingdom's HM Revenue and Customs and the Australian Tax Office) have begun to encourage the use of XBRL for corporate tax returns.

As described in more detail in Appendix A, administrative data are currently used in the United States mainly to create the business registers of the Census Bureau and the Bureau of Labor Statistics (BLS). Tax records from the Internal Revenue Service (IRS) form the basis of the Census business register, and ES-202 forms from state unemployment agencies forming the basis of the BLS Business List. Similarly, in the United Kingdom, administrative data from "pay as you earn" employee and value-added taxes form the starting point for the business register. In some European countries, especially in Scandinavia, essentially all interactions of individuals or businesses with the government generate administrative data that are maintained and available for analysis. One reason the United States has two different business registers is that the use of IRS data is extremely limited for confidentiality reasons. The advantages and disadvantages of a unified business registry are discussed below.

In addition to making use of administrative data, burden management should include methods for distributing survey response requirements across firms evenly, to the extent it can be done without damaging the representativeness of the survey data. Two equivalent firms should, over time and surveys, be asked to carry a similar burden—acceptable burden should be defined to be proportional to the size of the firm. Cooperation of most or all of the largest enterprises is required in nearly all federal surveys. Small to

medium-size firms are only sampled, and it is this domain of firms where the greatest need for burden management exists. Given the theme of this report that a need exists to increase sampling of young and small firms in business surveys, it is especially important to develop mechanisms to help spread the burden across those firms.

Two different approaches can achieve this ideal level of burden sharing, and both should be examined by survey managers and statisticians. The first approach may be called the *PRN* system, which entails the assignment of a permanent random number to each firm (or establishment) in the register. The random number would be stored as a separate field and would be available permanently both for a given survey over time and to different surveys across agencies. The random number would be used for specifying probability samples of firms. Because the number is permanent, it would provide a basis for spreading the sampling across all firms in the population. The implication is that all firms would be asked to respond to a similar number of surveys, leveling the response burden across firms. Some firms would not be asked to respond to a disproportionately large number of surveys, which is what could hypothetically happen today. The coordination or management of reporting burden according to a PRN system has been implemented in Sweden and is described by Ohlsson (1995).

The second approach may be called the *burden budget system*. If X_i is the size of the i-th firm, then define a burden budget B_i proportional to X_i, for each firm in the population. The burden budget might be expressed in units of hours, completed survey questionnaires, or the like. The burden budget would represent the firm's total obligation to respond over a defined period of time, such as five years. Every five years, the burden budget might be reset to its original value. The business register would record the specific surveys for which each firm was selected to participate, both over time and across agencies. With each survey in which the firm participated, the cumulative burden budget would be reduced by a measure of its participation. At any given time, B_i would represent the firm's remaining burden in the five-year period. In sampling for a new survey, the firm would be selected with a probability proportional to, or otherwise positively related to, B_i. Thus, as the firm fulfills its assigned reporting obligation through participation in surveys, its cumulative B_i declines and its probability of selection in future surveys declines. In this way, firms that have not participated eventually do participate because their cumulative B_i has not declined and, correspondingly, their probabilities of selection in future surveys increase. A simple system of this type operates in the United Kingdom for small firms. Participation in a survey in a given year earns very small firms a guarantee that they will not be asked to participate in another survey for a set period—three years in the UK example (Office for National Statistics, 2005). A

similar regime is also being considered for adoption in Denmark and the Netherlands.

It is apparent that a central business register and the information it contains are essential to driving the kind of burden management system described here. Thus, we turn now to describing the characteristics of the ideal business register, before moving on to more general aspects of the business data system.

3.2 DEFINING AND TRACKING BUSINESSES OVER TIME—THE BUSINESS REGISTER

3.2.1 Ideal Business Register Characteristics

The ideal business register is no doubt more easily described than created. Nonetheless, it is important to first consider the ideal, before turning to the possible, especially since what is feasible tomorrow will be different from what is feasible today. A business register must first be comprehensive; it must cover the entire business population of firms conducting business in the United States. The implication, then, is that the ideal register would include not only employer businesses, but also nonemployers, and would pick up firm births and deaths with very little lag. Although 100-percent inclusion rates are impossible to attain, the coverage must still be substantial and known. The business register should also indicate the enterprise structure of a firm, such that subsidiaries and multiple work sites are apparent, and mergers, acquisitions, and other status changes are reflected in a timely fashion. To maximize the value of a business register as a sampling frame, these data items must be included and should be available to all approved users of a register (including those positioned beyond the Census Bureau and BLS). In order to manage respondent burden, it is necessary to maintain a complete record of the sampling histories of firms and establishments. As noted above, this would include recording which survey samples the business has been selected into and what the reporting history was for that survey. Note that the complete contact information must also be included.

An additional set of items should ideally be linkable to these core business register items, with appropriate safeguards for confidentiality. First, given the importance of the ability to analyze the entrepreneurial process, indicators of both firm size and ownership type should be available, with owner demographics provided for privately held firms. Because small-domain estimates are important, detailed geocodes, industry codes, and product codes are essential. In order to link different aspects of the data with the main business register, a unique identifier is also needed (see below).

The ideal business list would also provide a unified sampling frame, in contrast to the overlapping, but not entirely consistent business lists currently maintained separately by the Census Bureau and BLS. A growing amount of time and effort is being spent comparing and attempting to reconcile these different lists. The goals of the major comparison project (described by Paul Hanczaryk of the Census Bureau and James Spletzer of BLS in National Research Council, 2006) are twofold: to understand the differences in the lists and to identify the strengths and weaknesses of each. There are indeed clear advantages and disadvantages to alternative administrative and survey data from BLS and the Census Bureau, respectively. Thus, "unified" should not be interpreted as drawing from a single administrative and survey source or agency. Rather, an ideal business list would integrate multiple sources so that a common registry is used for key business surveys underlying the National Income and Product Accounts (and related federal business statistics).

3.2.2 Unique Business Identifiers

The ideal business list would assign a single, unique number that would be used to identify each business in all data sources (including household surveys, for example, when listing the employer of record). This unique identifier would then give users the ability to link business registry information to data items from both employer and demographic surveys that provide data on worker characteristics, self-employment, and household-centered businesses. It would also allow the linking of business registry information to individual-level administrative records, such as those from the Social Security Administration and the Longitudinal Employer-Household Dynamics Program. Another benefit of a unique business identifier is its potential to help minimize respondent burden. By simply entering its one ID number, a business could update all of its information at once. In addition, creation of cross-walks between the unique business number and the current Unemployment Insurance (UI) number, along with data sharing by the state, would facilitate updating of the business list on a more timely (quarterly) basis.

Note that, as discussed earlier, what is meant by a business may differ by sector and can potentially change over time. Thus, it will be important for the register to preserve or develop indicators for the legal form and type of business entity. As part of the latter, registers should indicate franchise operations, multilevel marketing operations (e.g., Amway), joint ventures, special-purpose financial entities, and the like that are not captured by traditional "legal form of organization" indicators. It may even be necessary to allow for multiple definitions. One might also consider including additional information on ownership relations—for example, in cases in

which one firm owns a portion of another, or in which a firm is jointly owned by two others but neither exercises unilateral control.[4]

Use of a unique identifier also facilitates the linking of basic business register data to more in-depth firm information, thus enhancing confidentiality. That is, increasingly detailed information can be made available to qualified users, perhaps through the use of a relational database, while, for other uses, a more limited set of data items may be sufficient. Many kinds of users could legitimately claim a need to access information on the existence of a business, but additional safeguards must be in place for users granted access to more detailed information. Examples of more detailed information that could be linked to the business register include financial information (such as bank loans data, which might be bought from credit bureaus), as well as detailed and consistent product descriptions that may be in widespread use in the private sector. Examples of the latter include universal product codes for retail firms, national drug codes for pharmaceuticals, and insurance reimbursement codes for the health industry.

Finally, in certain sectors for which the government has special data sources, such as farming, railroads, and nonprofit and public organizations, the unique identifier could be used to link these data. Other governments have already moved forward with respect to this ideal. For example, in the United Kingdom, a major review by the Treasury to identify ways to reduce the administrative cost of regulation concluded that a single business identifier, in combination with data sharing across agencies, could meet this goal. Data sharing is key to reducing duplicate data requests (Hampton, 2005). It can also help streamline the administrative agencies themselves and lead to increased efficiency (O'Donnell, 2004).

In the United States, reconciliation of the major business registers generated by the UI system and the Census Bureau would be a major step forward. To the extent that reconciliation of establishment lists can take place as soon as data become available, the ability to rapidly and continuously update sample forms would be improved. The presence of unique geocode identifiers would also help facilitate data integration across sources.

3.2.3 Effective Data Sharing

By this point, it is clear that creation of the ideal business register implies a greater degree of data sharing among statistical agencies than is

[4]The statistical agencies (in the United States and abroad) have done quite a bit of work on the issue of determining ownership structures of businesses—see Armington (2004).

currently in place.[5] We have already cited the business list reconciliation example. Because there are advantages and disadvantages to both the BLS and Census Bureau systems—and thus reconciliation should take place downstream to minimize loss of information—a need for data sharing between these two agencies is implied. For example, data on sole proprietors, which originate from Schedule C filings to which only the Census Bureau currently has access, would need to be included. Even without worrying about reconciliation per se, a unified register implies data sharing. Note, though, that since the Schedule C data will include all individuals, including those whose activities can more accurately be described as a hobby than as a business, it is likely that some revenue threshold would need to be imposed before a sole proprietor is actually placed on the list.

Not every piece of information on the business register needs to be accessible to every agency, but a much broader range of users should be allowed than is currently the case. The idea of different agencies having access to the business register, with a well-defined set of rights (in terms of which fields are available) and responsibilities (in terms of confidentiality), is similar to a point made for the United Kingdom (Hampton, 2005). The recommendation there was that various agencies should be given access to only those fields in the database that were necessary for their regulatory efforts. The central database, along with data sharing, would play a large part in reducing the administrative burden on UK companies.

In addition to data sharing among agencies, the basic elements of the employer section of the business register should also be made accessible to all qualified agency and research users to serve as a sampling frame for their surveys.[6] Researcher access to some confidential data is already allowed in

[5] "Data sharing" is the exchange of information collected from businesses and individuals or reported to the IRS in identifiable form for statistical purposes. Identifiable form means information "that permits the identity of the respondent to whom the information applies to be reasonably inferred by either direct or indirect means." Statistical purposes involve "the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups." They also include methods and procedures related to the "collection, compilation, processing, or analysis" of data about these groups and the development of related "measurement methods, models, statistical classifications, or sampling frames" (National Research Council, 2006, p. 56).

[6] Nonagency use of a business list should be tightly monitored. One danger is that, if there is a business register that can be used as a master sampling frame then, potentially, more surveys will be launched, particularly if the cost for using the frame is trivial. This could lead to the unacceptable consequence of increasing burden placed on business respondents. Our primary concern here is with agencies' programs that currently must construct frames (or purchase them from Dun & Bradstreet) because they do not have access to BLS or the Census Bureau lists.

BOX 3-1
The UK Business Data Linking Project

The ONS collects large amounts of business microdata. Their Business Data Linking Project provides access to the data via its secure "microdata lab," where academic researchers can carry out statistical analyses. These data are confidential and access is tightly restricted:

- Only researchers fully employed at bona fide academic or charitable research institutes, or civil servants, may have access. There is no facility at the moment for PhD students.
- The employer is required to sign an agreement taking collective responsibility for the actions of all its researchers. Researchers are required to agree to standard secondment contract terms. There is no access without signed agreements.
- Projects must be of academic value and demonstrate (a) a clear interest for ONS in the results, and (b) a specific need for the data sets requested.
- Access is granted only through the secure microdata lab onsite at ONS.

Researchers must specify which data set(s) they want to use, why they want to use it, and why the data cannot be found elsewhere. In some cases, additional data sets may be linked by researchers.

the United States through the safeguards provided for via research data centers. Other countries have programs through which researchers can access data more broadly. For example, in the United Kingdom, the Business Data Linking Project (see Box 3-1) allows access in a secure setting to "safe people" who have been vetted by the Office of National Statistics (ONS).[7] More broadly, ONS allows access to a broad range of unpublished microdata to researchers upon application.[8]

3.3 IDEAL DATA COLLECTION CHARACTERISTICS

3.3.1 Contents of the Ideal Business Data System

Not only must the ideal business data system be capable of producing accurate and timely aggregate statistics on income and output, employ-

[7]See http://www.statistics.gov.uk/about/bdl/ for more details about the program.

[8]Much of this has resulted from the Deregulation and Contracting Out Act of 1994, which has been applied to the Statistics of Trade Act of 1947 and others. See http://www.statistics.gov.uk/about/NS_ONS/ONS_microdata_releases.asp for further details, including a list of data releases.

ment, investment, prices, and productivity, but it also should allow research that requires use of microdata. In particular, as noted above, individual-level business data are important for studying such areas as productivity growth, firm entry and exit, the role of young and small businesses in fueling economic growth, the characteristics of business owners, and the interactions between large and small businesses. A system capable of such analyses must thus collect a range of key data items beyond the basic information contained in the business register, although the register is certainly one important input. Measurement of transitions (e.g., mergers, acquisitions, and spin-offs) allows researchers to identify such things as expanding employment areas, structural changes in the economy, and business cycle turning points.

Just as important as the capacity to track transitions for existing businesses is the ability to detect the birth and growth of new firms. Maintaining ownership type and owner demographics on the business register is just the first step. The data system must also facilitate measurement of gross flows from employee to self-employed categories. To best understand these nascent entrepreneurs, additional information is necessary. First, it would be useful to be able to identify the location of self-employment—specifically, whether the business activity takes place at home or somewhere else. Data capturing time use in different business activities are also essential.

In order to understand the underlying sources of growth in the economy, data are needed on a range of business attributes that may be linked to performance. Tangible capital is one important area for which information is sparse, particularly for small firms. Becker et al. (2006) have shown that, in the Annual Capital Expenditures Survey, entrants and younger firms are underrepresented, even though the evidence suggests that capital expenditures in the start-up process of businesses are high. If true, the current survey is missing an important, and nonrandom, component of capital expenditures for U.S. businesses.

For new and growing firms, knowing more about financing is of particular interest. A business data system would, therefore, ideally include financial and balance sheet information, including equity, debt financing, and venture capital financing; measures of capital stocks and investments would also be useful. Important expenditures include investments in physical capital and also in technology and research and development (R&D), human capital, and organizational capital. Corrado, Hulten, and Sichel (2006) argue that better measurement is needed of three types of intangible capital: computerized information, knowledge acquired through scientific R&D and creative activities, and economic competencies. The ideal business data system would allow users to calculate intangible capital figures for such categories as stock of education and training in a firm and new investments in human capital and organizational capital. We emphasize

this as an ideal recognizing that, in practice, collecting information of this kind, especially from smaller firms, would be a challenge.

Black and Lynch (2005) argue that a firm's organizational capital—broadly categorized as workforce training, employee voice, and work design—contributes in significant ways to its productive capacity. The authors conclude that calculating the stock of training in a firm is difficult; however, even capturing measures of new investments in training would be a significant improvement. At the moment, U.S. statistical agencies collect no ongoing measures of the amount of training workers acquire year over year. This is in sharp contrast to the European Union, which includes training questions in all member countries' annual household surveys. Some of these kinds of questions could be incorporated into the Current Population Survey as well as most business surveys.

Other measures of organizational practices, such as the percentage of workers meeting on a regular basis to discuss workplace issues, unionization, layers of management, benchmarking usage, and the existence of incentive-based compensation could, in principle, also be gathered at the establishment level. Doing so would create minimal respondent burden since, as Black and Lynch recommend, training and compensation data need be collected only on an annual basis. Other components of organizational capital, such as employee voice and work design, could be collected on a less frequent basis; every other year is likely to be sufficient since these practices do not change with high frequency. With minimal respondent burden, policy makers and businesses would then have the ability to understand how these dimensions of intangible capital impact the productive capacity of individual firms and the economy more generally. In 2004, Eurostat began collecting information on organizational innovation in its Community Innovation Survey. Eurostat broadly defines organizational innovation as changes in firm structure or management methods that are intended to improve a firm's use of knowledge, the quality of goods and services, or the efficiency of work flows. The survey then operationalizes this definition with a range of more specific questions.

A full understanding of business success (and failure) requires data not only on these important inputs, but also on outputs, performance, and the market environment. Thus, the ideal business data system would go beyond the detailed industry and product codes on the business register and include such things as productivity measures, profit levels, prices, and even patent information. In addition, it is important to have information on business relationships, and to be able to link a given firm to its suppliers (or whom it supplies) and competitors. Related to this is the need for indicators identifying a franchise or a licensee, or a spin-off from other firms. Clearly, this broad list of information comprising the ideal business data system need not be available for all firms. Rather, the system would make appropriate

use of sampling and surveys. The following sections discuss more fully whom to survey, for how long, and how often.

3.3.2 Whom to Survey

Many of the interesting questions related to business dynamics are addressed by statistical estimates that measure relative changes in activities. On the one hand, a focus on dynamics forces attention to the act of creation of new business entities, a moment at which fundamental changes take place. Since the start of entrepreneurial activities is often limited in scope and magnitude, high growth rates are possible. On the other hand, many of these new entities die quickly. To understand the dynamics of business entities, nascent units must be included.

A critical question, addressed in Chapter 2, is, at what moment does a new business entity begin? Is it when the founder first thinks of the business opportunity? Is it when the founder begins active planning to launch the activity? Is it when the founder creates legal entities? Is it when the entity first takes actions toward producing goods or services? The business lists currently used in the federal government require longer and perhaps more elaborate business activities for inclusion in the target population than would be implied above. Hence, reliance on the current sampling frames may miss much of the economic activity of key interest in business dynamics.

A fully developed data system may require a new blending of demographic and economic surveys. Demographic surveys are typically defined as surveys of households and persons; sampling frames consist of addresses and other indicators of housing units in which people reside. Self-reports by the households in answering questions about employment, health, victimization, etc., form the basis of demographic statistics. It may also be possible to identify and measure the volume of nascent business entities through surveys of persons who are in the process of founding them. Samples of these persons could come from demographic surveys now ongoing, perhaps through question modules that ask "screener" questions about the entrepreneurial activities.[9] If a sample person reports the requisite activities, then he or she would be eligible for further measurement over time to track those activities. While, in an ideal data system, one would like to have this information, identifying start-ups by household (the majority of which do

[9]As described in Chapter 2, substantial work has been done in association with the Panel Study of Entrepreneurial Dynamics to develop cost-effective procedures for identifying entrepreneurs using household surveys.

not include individuals engaged in entrepreneurial activity), then subsequently linking the information to business identifiers to follow firms from genesis forward is not a trivial task.

Use of demographic surveys as a screening tool for new businesses must also acknowledge the problem of multiplicities. Multiplicities exist in sampling processes when the target population unit (in this case, a new business) might possibly be linked to several frame elements (in this case, persons within households). When two or more persons are engaged in the joint creation of a new business entity, then both of them could report it in a demographic survey. Thus, entities with multiple founders would be overrepresented in demographic surveys of persons. Such multiplicities would have to be measured as part of the screening process in order to create selection weights useful for inference to the population of business startups. A further source of multiplicities is the fact that some business entities reported by demographic survey respondents will already be listed in the existing sampling frames; others will not. This issue, too, would have to be addressed with selection weights.

Another issue tied to using demographic surveys is the loss of unambiguous and objective criteria for eligibility (as can be the case with Current Employment Statistics' use of UI filing). Self-reports by household members about their business activity would create the basis for eligibility into the sampling process. Careful development of screening questions to ensure that accurate reports of eligibility could be obtained would be essential.

3.3.3 How to Allocate the Sample of Business Entities

It is simple to show that the precision of estimates of population totals in business surveys is maximized when the sample of business entities gives very high probabilities of selection to very large firms and small probabilities of selection to very small entities. Typically, high-volume firms are sampled with certainty (i.e., they always appear in the survey sample). This allocation of the sample ensures that large portions of economic activity are represented in the sample.

With a focus on business dynamics, however, when the interest is in relative changes in business activity, then alternative allocations of the sample become more attractive. Allocation of the sample proportionate to some function of change in business activity is reasonable. Depending on the estimates of interest, preferred sample allocation might place more emphasis on small, quickly changing new businesses.

Since these two purposes—estimating total volume and estimating rates of relative change—suggest different allocations of the sample, some consideration might be given to a new data collection vehicle of business dynamics that could be viewed as a supplemental survey for the current

economic measurement systems. This survey would allocate the sample disproportionately to new business starts, those strata of economic activity that form only small parts of current economic surveys. The data from this new survey, perhaps called the Survey of New Business Dynamics, could be combined with that of other surveys to provide full estimates of relative changes in key activities.

Another sample allocation issue concerns the sector of economic activity to be covered. This is best illustrated by considering e-commerce, a sector in which intensely dynamic business activity takes place. A new focus on business dynamics would require much more attention to such sectors that pose new survey measurement problems. Units in the e-commerce sector need to be identifiable with well-defined geographic locations—yet entities exist in cyberspace. The physical location of the server infrastructure can change instantaneously, which makes it difficult to link people and places to business activity. Although the federal statistical agencies have programs addressing e-commerce, studies of business dynamics would benefit from increased investment in survey methodology for technology sectors. Similarly, the activities of nonprofit entities are highly dynamic over time, so that sector could also be part of a new program in business dynamics.

A further sample allocation feature that must be addressed is the nature of the longitudinal structure to be carried out in the survey. To maximize the precision of estimated changes in the volume of economic activity, it is common to measure sample units on a frequency that is related to the historic rates of change of the phenomena measured. Phenomena that exhibit relatively rapid change (e.g., employment and unemployment) should be, and typically are, measured more frequently than phenomena that are less volatile (e.g., state government tax revenues). A statistical system that measures highly volatile attributes frequently and less volatile attributes less frequently could be constructed through a set of questionnaires that have flexible modules that are used at different intervals over time. Such a flexible design can reduce the burden of measurement on sample businesses and can be used to gather information at appropriate intervals.

3.3.4 How Long Should a Sample Business Be Measured?

Many of the current economic surveys have longitudinal designs; that is, a sample business enters the survey and is repeatedly measured over time. While such a design theoretically permits the analysis of micro-level changes in specific business entities (e.g., gross changes in size over time within a sector), the designs are traditionally not exploited for that purpose. Instead, because many of the surveys are used to produce estimates of

change in totals over adjacent time periods, precision is enhanced by having overlap of sample businesses.

Studies of business dynamics require following the same business over time, measuring similar phenomena repeatedly, and assembling longitudinal data on each sample business that can be used to examine precursors to change, what types of businesses experience different types of change, and which types of characteristics of the birth process lead to more or less stable entities. Such purposes require that more attention be paid to following time-based linking rules than is true for most current business surveys. Several problems arise and need further study: How are mergers and acquisitions to be handled in the study of business dynamics? How are spin-off businesses from the original entity to be handled? When is a new business de facto the death of a sample business, and when is it a new component of the original sample business?[10] In demographic surveys, there exist various following rules, depending on the purposes of the longitudinal survey, some of which permit the inclusion into later waves of the panel persons who join families originally sampled into the survey. For example, the Panel Study of Income Dynamics, which started with a sample of families in 1968, has continued to follow not only those families, but also the new families formed by any member of the original family.[11] Such designs permit more complete understanding of the temporal dynamics of families and households. Similar following rules could also be used in carrying out business surveys.

Longitudinal surveys are used to produce estimates of the population at any one moment in time *and* changes in the population as defined at earlier points. Designs with such goals are often *rotating panel surveys* that, at any one point in time, sample new units and rotate out sample units that have been measured over prior multiple waves. Such designs have the advantage over *fixed panel surveys* (which draw a sample at one time point and follow it for long periods of time without any refreshment of the sample) by reducing the respondent burden and providing both micro-level change estimates and cross-sectional estimates.

A key issue in rotating panel designs is the number of waves of measurement for each sample unit. In line with the logic above, studies of

[10]Beyond the work done in the United States (discussed throughout this report), approaches for dealing with problems related to mergers and acquisitions, and births and deaths, have also been recommended by the Organisation for Economic Co-operation and Development, Eurostat, and others internationally—see Pilat (2002) for an overview of methods used in a range of countries (and for further references).

[11]See http://psidonline.isr.umich.edu/ for more information on the survey and how it follows family members over time.

business dynamics might follow units subject to high likelihoods of change for longer periods and units that are unlikely to undergo major changes only briefly. Determining the strata for long follow-up and short follow-up would require some study for each outcome variable selected.

Finally, a separate issue is the length of the interwave interval—how often should a sample unit be measured? If life expectancy is low for new businesses, frequent measurement is needed to acquire observations about life-cycle processes. There are many attributes of new businesses for which information is essential (e.g., payroll, employment counts). To reduce the reporting burden of high-frequency measurement, new alliances between the federal statistical agencies and payroll processing firms might be considered; use of scanner data for sales volume estimates might also be studied. Other variables may not require such frequent measurement to be useful (e.g., use of new technologies). Clearly, measurement of business dynamics would benefit if the method of record construction permitted data to be acquired from units in a temporally flexible manner.

4

Limitations of the Current Data System for Measuring Business Dynamics

The U.S. statistical agencies provide a wealth of information about the activities of U.S. businesses.[1] Moreover, in an era of tight budgets, they have continued to innovate and respond to measurement needs for a rapidly changing economy. Despite these successes, there remain a number of large data gaps in U.S. business statistics. A key deficiency relates to data integration and coordination. The system of economic accounts is balkanized given that the underlying source data are obtained by multiple agencies. As emphasized in a recent volume on the "architecture" of the national accounts (Jorgenson, Landefeld, and Nordhaus, 2006), considerable gains could be realized through better integration of data—both administrative and survey based—some of which would require more extensive data sharing between the Bureau of Economic Analysis (BEA), the Bureau of Labor Statistics (BLS), and the Census Bureau.

One way to evaluate systemic data gaps is to ask whether it is feasible to go from the aggregate statistics to the firm level in a consistent manner. The feasibility of disaggregating data is of interest for a number of reasons. It is useful as a tool to help identify data gaps, since the inability to drill down to the micro level may reveal situations in which there are limited or no source data underlying an aggregate statistic. An example is the measurement of capital expenditures by detailed asset type and by industry.

[1]Appendix A provides an overview of the major federal business data sources.

Such statistics, produced by BEA, are used frequently to gauge which industries are heavy users of advanced technologies. However, underlying source data are extremely sparse and very strong assumptions are required to produce these statistics (see Becker et al., 2006). More generally, the ability to disaggregate data down to the firm level has practical applications for both the measurement and interpretation of business statistics. For measurement, anomalies can be identified through scrutiny of the underlying source data. For interpretation, observed changes may reflect fundamental compositional changes in the mix of firms that are important for understanding the business cycle or secular trends in the economy.

It is possible to connect some key aggregate statistics, such as employment and payroll, to the underlying firm-level data, and recent programs at BLS and at the Census Bureau exploit this micro-macro link. However, for most key aggregate statistics, particularly those that require combining nominal values (such as nominal gross output) and prices, this is not feasible. The reason is that nominal activity measures for firms are collected by the Census Bureau, whereas nominal prices are collected by BLS; these data cannot be shared and thus they cannot be integrated at the firm level. Even with data sharing, coordination and integration of the administrative and survey data would be required to permit such drilling down from aggregate statistics. In short, for key national statistics such as real value added, real capital expenditures, or real productivity growth of a sector, it is impossible to connect back to firm-level statistics. Again, this shortcoming reflects data gaps as well as lack of data integration.

This report is primarily about the importance of measuring changes in businesses. As has been emphasized throughout, the existing system of accounts and underlying source data focus on measuring the level and cross-sectional variation in the levels of activity. The neglect of business dynamics and the role of young and small businesses in accounting for growth are related to the data gap problems discussed above. In order to capture the contribution of young and small businesses, the underlying administrative data tracking businesses must include them in a consistent and coordinated manner across the statistical agencies. Moreover, administrative data, while a powerful source of information on U.S. businesses, do not include a number of important firm characteristics that must therefore be collected by surveys. The neglect of young and small businesses by the existing system implies that a number of these key measures are not collected in a representative manner.

To explore these data gaps for measuring business dynamics, this chapter reviews the nature of data coverage of U.S. businesses past and present. Since administrative data sources can be used to track business dynamics, and in particular young and small businesses, we then turn our attention to data gaps in the business registers. However, since administrative data must

be supplemented with survey data, we also describe weaknesses in existing surveys. Finally, since a significant part of the data gap for measuring both levels and growth of economic activity reflects a lack of data sharing and data access, we also discuss these issues.

4.1 DATA COVERAGE OF YOUNG AND SMALL BUSINESSES

For much of its history, business data coverage in the United States was by and large limited to the agriculture and manufacturing sectors, reflecting where much of the economic activity in the country was located. Some of the earliest efforts by private-sector firms to count and characterize firms included those by R.G. Dun and Company, which produced credit reports on firms in the 1840s, with information on owners, ownership structure, firm performance, and financial transactions. From 1842 to 1890, the company amassed records on some 1.2 million firms (Madison, 1975). Dun & Bradstreet data extend back sporadically to the 1870s for some sectors of the economy.

On the government side, various censuses of manufacturing were conducted, some at the state level, prior to 1900. The 1810 Census of the Population was the first to include questions—albeit a highly limited set—on manufacturing. The Office of Business Economics, the precursor of BEA, began producing data on the number of firms by major industry in 1929 and continued to do so through 1963, at which time the program was discontinued due at least in part to concerns about data quality. By the mid-1950s, the quinquennial economic census was developed to the point that the Census Bureau's Enterprise Statistics Program was able to produce statistics on firm size for selected industries. However, even with additional input from the bureau's County Business Patterns—begun in 1946 but not available annually until 1964—data production was sporadic and incomplete, typically released with four- or five-year lags (Armington, 2004, p. 2).

It was not until the development of comprehensive business lists that the statistical agencies began generating high-quality descriptive data on something approaching the universe of U.S. employer businesses. Historically, most data collection by federal statistical agencies has been oriented toward a cross-sectional view of business activity. While progress has been made more recently at the statistical agencies to develop and improve statistics on business dynamics, there are still important topics that cannot be adequately explored due to data deficiencies.[2]

Figure 4-1 summarizes the distribution of business data coverage, mapping data sources in terms of the business and labor populations they cover

[2]See Appendix A for a review of business data sources relevant to business dynamics.

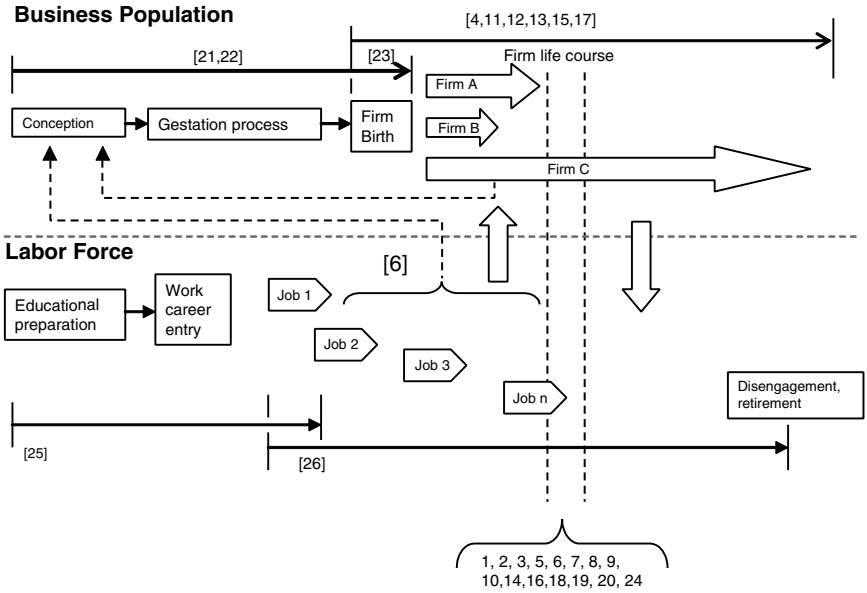

FIGURE 4-1 Stylized depiction of current data coverage.

Key to Numbered Data Sets

1. BLS, Business Establishment List
2. BLS, Quarterly Census of Employment and Wages
3. BLS, Current Employment Statistics
4. BLS, Business Employment Dynamics
5. BLS, American Time Use Survey
6. BLS-Census: Current Population Surveys
7. U.S. Census Business Registrar
8. U.S. Census Company Organization Survey
9. U.S. Census, Economic Census
10. U.S. Census Survey of Business Owners
11. U.S. Census Longitudinal Business Database
12. U.S. Census Integrated Longitudinal Business Database
13. U.S. Census Longitudinal Employer-Household Dynamics
14. Dun & Bradstreet Duns Market Identifier file
15. NSF [U.S. Census] Longitudinal Research Database
16. SBA Statistics of U.S. Businesses
17. Business Information Tracing Series [BITS]
18. FRB Survey of Small Business Finances
19. IRS Survey of Income
20. Standard & Poor's Compustat
21. Kauffman Foundation Panel Study of Entrepreneurial Dynamics [University of Michigan]
22. Kauffman Foundation and Others: The Global Entrepreneurship Monitor
23. Kauffman Firm Survey [Mathematica]
24. Kauffman Financial and Business Databases
25. National Longitudinal Survey of Youth [BLS, conducted by Ohio State/NORC]
26. Panel Study of Income Dynamics [University of Michigan]

along a life-cycle continuum. The top portion of the figure categorizes efforts for which the unit of analysis is the business entity; the lower portion emphasizes individuals participating in the labor force. Dynamic processes associated with each unit of analysis are depicted by the horizontal arrows. As represented by block arrows labeled Firm A, Firm B, and Firm C, firms are developed and implemented, they grow and shrink and, at some point, many deactivate. Similarly, as represented by block arrows Job 1, Job 2, Job 3, etc., individuals complete their schooling, enter the labor force, take a series of jobs, and eventually withdraw and retire. Some members of the labor force become involved in creating new firms, represented by the dashed line from the job arrows to the block labeled "conception." Some new firms are created by existing firms, represented by the dashed line from "Firm C" to the block labeled "conception." The considerable cross-level activity is represented by vertical arrows.

The numbers in brackets refer to 26 different current data collection activities identified by the panel. Sixteen of these data collection efforts [1, 2, 3, 5, 6, 7, 8, 9, 10, 14, 16, 18, 19, 20, 22, 24] are designed to provide time-specific estimates, suitable for cross-sectional or cross-temporal assessments, but they do not allow for tracking the same units of analysis over time. Six [4, 11, 12, 13, 15, 17] allow for tracking of firm changes once they are incorporated into official business registries. Two [25, 26] capture individuals at different stages of their life-course work career. Two [6, 21] provide information related to individuals' efforts to create new firms. One [6] provides evidence of some efforts to create a self-employment option; the other [21] identifies and tracks individuals as they enter into the firm creation process (this data source is based on a one-time screening of a cohort identified in 2005). Finally, one specialized project [23] is designed to track the early years of the firm creation process.

The salient point here is that business data systems operated by the U.S. statistical agencies are designed primarily to capture levels of rather than changes in economic activity. The vast majority of data sources used in the construction of economic statistics are concentrated around mature, stable businesses and static components of the workforce. There is little coverage of early and late life-cycle dynamics and a paucity of information on business creation and start-up phase processes that take place before businesses are incorporated into official registries.

4.1.1 Business Register Consistency and Coverage

In assessing the current system for measuring business dynamics, it makes sense to begin at its core, with the business registries. In contrast to many other countries (see Box 4-1), there are four major business registers in the United States that provide wide-scale coverage of both publicly held

BOX 4-1
Alternative Models:
Building Registers Primarily from Administrative Data

Upon systematic review (see Appendix A), it is striking how many good business surveys are conducted in the United States. The problem is that in only a few cases can information be merged across sources. This makes it expensive to collect the data and also creates a substantial burden for respondents.

In contrast, some European countries now collect their census data on persons and businesses via a number of administrative registers. In these systems, individuals have identification numbers assigned at birth or, for persons not born in the country, when they are granted a work permit for the first time. This is similar to the U.S. Social Security numbering system. The register of identification numbers is kept up to date through several sources, again depending on the country. In most of Scandinavia, a housing register provides the core structure of the person registers, as all inhabitants are allocated to an address and to a particular house or apartment. Whenever people change addresses, they must inform the agency in charge of the housing register. All births are recorded in the birth register; there are other kinds of registers—e.g., tax, education, health, criminal—as well. In Denmark, administrative registers must be authorized by Parliament and statistical registers by Statistics Denmark; the conditions for data use have to be codified in the act allowing the collection of data.

Similarly, each business has an ID (similar to the employer identification number or EIN in the U.S. system) which also serves as the value-added tax (VAT) number. This number is given to a firm at the time of registration; all agencies collecting data use the same ID numbers, with the result that the register is constantly kept up to date with basic information on all citizens and business entities. Registering a business is required if it is incorporated; the business also needs a number in order to collect and deduct VAT, or if the owner wants his income to be taxed as a business. Since business tax is lower than personal income tax, owners have a clear incentive to register at an early stage. If the business has employees, it must register in order to collect taxes from them. All these registrations are coordinated through one single business ID number. Furthermore, each geographical workplace (establishment) has a register number that is linked to the business ID number.

Information on businesses and citizens is updated every time they have to report payment or withholding for persons. Every January, businesses must regis-

and privately held businesses; three are maintained by government agencies—the Internal Revenue Service (IRS), BLS, and the Census Bureau—and one is a private database produced by Dun & Bradstreet. The business registers at the Census Bureau and BLS are the primary sources from which business statistics on firm and establishment dynamics are generated. The main programs on business dynamics in the United States—Business Employment Dynamics (BED) and the Statistics of U.S. Businesses (SUSB) and related Longitudinal Business Data (LBD)—are constructed from the microdata on establishments in these files.

ter income earned by each employee—this is similar to the U.S. form W2. At the same time, they must indicate who was employed on a specific day (typically in November). Furthermore, for firms that maintain multiple work sites, employers indicate in the same form the physical location at which each employee worked throughout the year.

Government agencies are allowed to share data only to a limited extent and, even then, it must be permitted in the act that authorizes collection of specific data. However, Statistics Denmark can by law request all data, including the person and firm IDs, from each collecting agency, as well as from private sources. Statistics Denmark is thereby allowed to fully exploit the common ID numbering system and use all data in the production of statistics and for research. Trust and reliance in the system are so broad that a standard census has not been conducted since 1970. One of the conditions of use by Statistics Denmark is that no individual or firm can be identified. Researchers desiring access to confidential microdata that are allowed work inside the agency are subject to the same nondisclosure requirements as regular employees. For the past five years or so, it has also been possible for researchers to obtain admission to approved data sets from their own workplaces through the Internet (http://www.dst.dk/HomeUK/ForSale/Research.aspx).

Although the approaches described here may be more problematic to implement for an economy as large and complex as the United States (e.g., where unregistered and illegal employment is perhaps more widespread), universal coverage registers create interesting possibilities that do not exist with other systems. First, survey samples of persons and businesses can be drawn from the universal registers. This increases the quality of sampling frames and saves respondents time because they do not have to provide information already embedded in the registers (number of employees, industry, gender, age, education, children, etc.). Second, survey response data can be easily merged with register information. For example, responses from a survey of job and life satisfaction, in which only one or a few employees in each firm are sampled, have been linked with register information on all the coworkers in the firm. This has generated information on how the composition of workers, their wages, and the hierarchical structure of the firm may affect responses from the sampled person. Another example is analysis of the relation between human capital formation of employees and the bottom line of firms. A different class of examples can be found in medical research, where groups of patients can be traced back in time with controls for work- or residence-related exposure applied. An example here is work on the long-term effects of occupational hazards.

The BLS business register, constructed primarily using data that originate from the Quarterly Census of Employment and Wages (QCEW) program,[3] is designed to produce frames that are optimized for survey mea-

[3]The QCEW is augmented with survey data—the Annual Refiling Survey and the Multiple Worksite Reports are particularly important. The former is used to obtain updated information on business names, addresses, industry codes, and contact information. The latter is used to improve the accuracy of establishment-level employment and wage figures for multiunit firms operating within a state.

surement at relatively high frequencies. The QCEW program is administered by the state Unemployment Insurance (UI) programs. For their register, the Census Bureau relies on the federal tax system, the Company Organization Survey (COS), and the quinquennial Economic Censuses to maintain a frame optimized for periodic survey measurements at relatively low frequencies. The end result is two fully maintained establishment registries with very substantial overlap (probably in excess of 95 percent of establishments occur on both lists) and two programs for creating business-based statistics. The Census Bureau's measurements are focused on the outputs and the nonemployment inputs while BLS's measurements are focused on the employment input. Consequently, the national statistical products do not reveal the extent to which the agencies' establishment list inputs contain the same entities.

Business list comparison projects conducted cooperatively between the Census Bureau and BLS indicate that, in some areas, the business lists of the two agencies do not match up well. Even after controlling for differences in scope and coverage between the two registers, the aggregate employment figures (published in *County Business Patterns* (CBP) and the QCEW) vary significantly. For example, in 2001, the CBP data show a total employment of about 115 million, while the QCEW figure was not quite 109 million—a difference of about 5.5 percent.[4] One difference in scope between these two business lists is that the Census business register includes nonemployer businesses while the QCEW does not (however, statistics on nonemployer businesses are published separately from the CBP). Inconsistencies in the registers and associated aggregate, industry, and firm-level statistics may stem in part from the differences in the treatment and coverage of business dynamics and young and small businesses. It is inherently difficult to track young and small businesses—they tend to increase or decrease in size rapidly—so they are a moving target for measurement purposes.

Another dimension in which the business lists vary is in their identification of single versus multiunit status. In BLS's Multiple Worksite Reports (MWR) system, multilocation employers with at least 10 employees in secondary locations are required (in some states) to break out their employment and payroll by individual establishment.[5] When a single-location firm expands to multiple locations, the "new" establishment will not be observed until it crosses the 10-employee threshold. In addition, since MWR

[4]BLS and the Census Bureau are working together to better understand the sources underlying this difference—at this point, we cannot infer whether one statistic is better or more accurate than the other.

[5]See http://www.bls.gov/cew/cewmrr00.htm.

is a state-based program, if the expansion occurs to a state in which the firm does not already have a presence (and if the firm does not use a unique national EIN), it is possible that it would not be captured as a multiestablishment birth but as a new firm in the other state. The Census Bureau business register program has excellent coverage of multiestablishment firms every five years but is uneven in intervening years, when the COS must be relied upon. The COS includes all employers with at least 250 employees but only a sample of companies smaller than that.

These differences in the processing of multiestablishment firms across the two registers result in comparability issues, especially with regard to time series movements in births and deaths, and to job creation and job destruction series. Comparisons made using the 2001 microdata indicate that there are about 309,000 cases in which the Census Bureau and BLS lists disagree in the single versus multiunit designation, and these businesses employ 21-22 million people.[6] Even for the 4.1 million businesses that both agencies agree are single establishments, BLS estimates a total employment level of 35 million, and the Census Bureau a total of 38 million. Analysis of the 2001 microdata indicate that 71 percent of these establishments match exactly on payroll figures, and 69 percent match exactly on employment figures. For multiunit EINs, a near match (those within a plus or minus 2.5 percent band) occurs 51 percent of the time on payroll and 39 percent of the time on employment. There is also significant mismatch in industry classifications at the micro level. For single-unit establishment firms found in both lists, less than 70 percent match at the 6-digit level of the North American Industry Classification System (NAICS). This level of disaggregation is often used to produce the national accounts, raising obvious concerns about the consistency of source data used to build that all-important set of economic statistics.

Differences in the way that business structure and single-unit/multiunit transitions are covered in the two major business lists lead to inconsistencies in estimates of important dynamic activities associated with births, deaths, and restructuring. The heavy reliance on the Economic Census for new information about business structure is reflected in the measures of dynamics derived from that source. The result is clusters of apparent activity at five-year intervals that is really cumulative over the period between the censuses. At the same time, reliance on state employment security agencies for information about business structure in the BLS list may result in state-to-state variation in the quality of business structure information.

[6]BLS and the Census Bureau have begun documenting results from comparison projects. See, for example, the summary of a presentation by Spletzer and Hanczaryk in National Research Council (2006); see also Foster et al. (2005).

There are also distinct differences across the registers in terms of how ownership links are recorded. BLS has reasonably complete accounting of ownership links within a state for employers reporting a single EIN. However, some large corporations with complex parent-subsidiary relationships may have multiple EINs, and ownership links for multistate firms are not comprehensive in these data. This makes it difficult to distinguish between de novo and experienced firm entry. The Census Bureau business register has high-quality ownership links for large companies that are maintained and developed through the COS and economic census programs. However, for smaller multiunit firms, details of ownership links across establishments degrade over time between censuses.

Inconsistencies in the coverage and accuracy of the business lists have direct implications for the reliability of key business statistics—from gross domestic product (GDP), to employment, to productivity and industrial production—that are all derived, at least in part, from business list data. A number of problems for downstream data users were discussed at a workshop on the topic of business data sharing.[7] Presenters from BEA described the growing statistical discrepancy that now exists between GDP and gross domestic income (GDI), two key aggregate measures of domestic output that should, in principle, agree. One source of the statistical discrepancy can be traced to the fact that GDP and GDI are calculated from different data sources, the former primarily from Census data and the latter primarily from BLS data (National Research Council, 2006, p. 15).

The most interesting (and problematic) discrepancies occur, however, at the industry level. For example, in 2002 the Census Bureau measure of current-dollar value added for the computer and electronic products sector was twice as large as that shown by BLS data, providing two contrasting pictures of factor productivity for the sector. Industry-based productivity studies typically rely on output data from the Census Bureau and input measures (industry employment) from BLS. These problems adversely affect the economic decision-making process. The ability of the Federal Reserve to conduct effective monetary policy could be diminished when productivity estimates are inaccurate because these data factor directly into assessments of inflation trends. Indeed, during the late 1990s, flaws in industry-level data resulted in slow recognition of productivity gains in technology and related industries, which directly affected assessments of the prospects for the nation's economic performance (National Research Council, 2006, pp. 29-30).

[7]For a summary of this workshop, see National Research Council (2006).

4.1.2 Register-Based Business Dynamics Programs

Longitudinal microdata are essential for measuring business dynamics—births and deaths, expansions and contractions, mergers and splits, relocations, ownership changes, and worker flows—the processes on which this report is focused. Because sources of longitudinal business microdata have historically been scarce, particularly for smaller and newer businesses, research progress on business dynamics and entrepreneurship has been hampered. Recently, however, data products have begun to emerge that promise to greatly enhance available information relevant to these topics. In theory, tracking businesses through time can be accommodated either by designing longitudinal surveys or by linking records from successive years. For the most part, the latter strategy—linking data, mainly across administrative and survey sources—has been used at the statistical agencies because of cost and respondent burden constraints. Because of their proximity to the business lists, the Census Bureau and BLS are the key players in this new area. Among the most promising data sets now coming on line are BLS's BED, and three Census Bureau efforts—the SUSB, produced jointly with the Small Business Administration (SBA); the LBD and its successor the Integrated Longitudinal Business Database (ILBD); and the Longitudinal Employer-Household Dynamics (LEHD) Program.

In addition to the establishment employment counts, the BED publishes quarterly statistics on job creation, broken out by expanding and opening establishments, and job destruction, broken out by contracting and closing establishments. Data are disaggregated by industry (sector level) and by firm size. The firm size data use tax EINs to aggregate across UI accounts. Development of the BED required BLS to overcome several methodological hurdles to create record linkages from EINs (in coordination with other information), which are imperfect for the purpose of producing firm-level statistics (see Okolie, 2004).[8] The SUSB program at the Census Bureau (partially supported by the SBA) uses the Census Bureau register to construct annual measures of job creation and job destruction disaggregated by opening, expanding, contracting, and closing establishments, along with respective establishment counts. The SUSB reports currently provide more geographic and industry detail than does the BED, breaking data down data by state, enterprise size, metropolitan statistical area, and 4-digit NAICS code (plans are in the works at BLS to produce statistics at finer levels of detail in these dimensions). However, the SUSB annual reports are issued with a significant lag (2 to 3 years) compared with the BED reports

[8]Spletzer et al. (2004) provide an overview of the BED program.

(which have a roughly 8-month lag). There are also differences in industry and type of employer coverage across these two data sets.

In addition to the SUSB, the Census Bureau also has two research databases constructed from the business register. The LBD is an establishment-level longitudinal data set that can be used to measure producer dynamics, including the birth and deaths of establishments and firms. Its coverage and properties are similar to the SUSB; however, it provides researchers with links to additional survey data. A new data effort at the Census Bureau is creating an ILBD, which includes data from both the employer universe (LBD) and the nonemployer universe. The objective of this project is to integrate these data sources by not only merging the data together, but also by developing micro-level links between the data on nonemployers and on employer firms. In this way, the transition of firms from nonemployer to employer status, and the reverse, could be tracked (Davis et al., 2006). Currently, very little is known regarding such transitions.

The LEHD is a relatively new microdata source being developed by the Census Bureau that integrates data on households and individuals with data on employers.[9] The program has created opportunities to conduct research on topics for which empirical analysis of confidential longitudinal linked employer-household microdata are required, such as research on low-wage workers and human capital and productivity. The LEHD, which has also been applied to investigate a range of other topics (e.g., outsourcing), illustrates the tip of the iceberg in terms of the information volume and detail that can be made available through data integration and the efficiency of the approach relative to developing new surveys.

An advantage of these newly developed research databases is that they offer the possibility to link additional information about firms or establishments into an analysis. At least every five years the Census Bureau carries out detailed surveys of these establishments. These surveys provide a much richer description of the activities of the establishment or firm. In contrast, the quality of both business registers suffers because they do not incorporate basic data on establishment output. The LBD, for instance, allows researchers to link the basic data on establishment and employment in the business register to richer data on inputs, outputs, characteristics of business owners, and other variables.

[9]The LEHD relies on BLS's QCEW to provide accurate detail on establishments (such as physical locations, industry codes, and employment numbers) and also for developing a correspondence between state UI account numbers and federal EINs for individual establishments.

4.2 GAPS IN DATA ON BUSINESS DYNAMICS AND ON SMALL, YOUNG, AND NASCENT FIRMS

As indicated throughout this report, the primary emphasis of the statistical system traditionally has been to produce accurate cross-sectional, highly aggregated statistics. One consequence of this strategy is that our ability to accurately measure the activity of small and younger producers, and their role in dynamic economic processes, is compromised. For this segment of the business population, a greater reliance on administrative records has developed over time.[10] Administrative data typically contain only a limited number of data fields such as payroll and employment.

Even with limitations imposed by administrative data, U.S. business statistics typically can be disaggregated along a number of dimensions—for example, by firm or establishment size. However, the federal statistical system produces very little information tabulated by firm or establishment age. A better understanding of business dynamics, industry evolution, the role of entry in markets, and the productivity impact of new firms requires data on business age. As illustrated by past efforts—for example, the 1939 and 1948 Censuses of Retail—and by recent studies using business microdata (Davis, Haltiwanger, and Schuh, 1996), it is certainly feasible to construct such statistics. A focus on publishing statistics by business age would also dovetail well with recent innovations in measuring producer dynamics, such as the BED, Quarterly Workforce Indicators, and LBD/SUSB programs. Since these programs rely on the accurate measurement of entry and exit of producers and the accurate tracking of existing producers over time, their statistical frames could be readily adapted to allow for the measurement of business age and the production of statistics disaggregated by business age. This involves no new collection of data but would require business registers to maintain information that, though relatively easy to construct at the establishment level, does require somewhat greater effort at the firm level, where adjustments for mergers and acquisitions must be incorporated into the company age data.

Besides placing greater emphasis on the production of statistics (both levels and dynamics) of small and young firms from existing data resources, more regularly collected information on the activity of these businesses needs to be collected to help fill several conspicuous gaps. For example, few data sources detail how young firms invest in research and development,

[10]A good example of this is the Census of Manufactures. Many statistics produced on small firms in the Census of Manufactures are imputed values based on administrative record data. Few small manufacturing firms in an economic census year actually receive forms. This clearly reduces costs and reduces respondent burden, but the result is that less is known about small and, thus, the majority of young firms.

capital equipment, or human capital. Recent studies (e.g., Becker et al., 2005) show that investment rates in physical capital by very young businesses are high, but these businesses are not covered in a representative manner by surveys of capital expenditures. There is a paucity of data on the activities of the nonemployer universe of firms, and even less is known about the transitions and interactions between the employer and nonemployer universes. Data on this transitional phase are essential for understanding the entrepreneurial process whereby firms evolve to the point at which workers are hired. Moreover, the level of activity in the nonemployer universe of firms is substantial. Census Bureau figures indicate that there are over 18 million nonemployer firms in the United States—roughly three times the number of employer firms. In general, little is known about the evolution of these firms (Davis et al., 2006).

While data on the universe of nonemployer firms (which is dominated by very small producers) actively engaged in business are sparse, even less is known about the firm-formation process in the preproduction phase. Before the presence of activity by a new entity can be detected in a business register, it must have either positive sales or employment (the two standard ways that administrative systems identify firms). In order to learn about entrepreneurial activities in the preproduction stage, a different data collection approach that surveys households or individuals (e.g., the Current Population Survey [CPS]) is most likely required. Fairlie (2006) uses the CPS questions about self-employment income to identify entrepreneurial activity, and the Kauffmann Foundation publishes an index of entrepreneurial activity based on these data. In order to measure entrepreneurial activities in the preproduction phase (before income is earned and tax returns are filed), one must ask specific questions about the topic in household surveys such as the CPS or perhaps BLS's American Time Use Survey.

Fairlie enumerated a list of advantages of data collected from households relative to that collected from businesses. Household data sources currently offer comparatively large sample sizes and long time series; more timely estimates of business ownership and entrepreneurship; built-in comparison groups of nonbusiness owners; the potential, when in panel form, for measuring entrepreneurship, business creation, transitions into and out of self-employment,[11] and for examining income growth (e.g., the National

[11]By linking the CPS files over time, longitudinal data can be created, which allows business creation to be examined. The Kauffman Index of Entrepreneurial Activity attempts to do just this: using matched data from the 1996-2004 CPS, all individuals ages 20-64 who do not own a business as their main job are identified in the first survey month. It is then determined whether these individuals own a business as their main job (15 or more hours typically worked per week) in the following survey month. Unfortunately, at the present time very little information is obtained about the nature of these new business activities.

Longitudinal Survey of Youth (NLSY), the Panel Study of Income Dynamics); the ability to include a wide range of variables measuring owner characteristics (e.g., race and gender, age, education, personal wealth); and the capacity to create detailed profiles of previous business ownership, work, employment and earnings experience (e.g., the NLSY), and to include questions on nonmonetary returns (e.g., job satisfaction) or on psychological factors related to entrepreneurship.

The disadvantages of household data have to do with limited information on such factors as business outcomes (e.g., sales and profits); employment (e.g., number, wages paid, health insurance, other benefits, gender and race composition); customers (e.g., other firms, consumers); revenues (e.g., from government contracts, international sales); financial inputs (e.g., start-up capital, types of financing, bank loans); family businesses and work experience in family businesses; business acquisition processes (e.g., franchise, inherited, purchased, family gift); innovation (e.g., patents, new service or product); and use of technology (e.g., computers, Internet).

The statistical agencies do have programs that target small firms (see Appendix A). A clear example is the Survey of Business Owners (SBO), last survey conducted in 2002.[12] The SBO surveys sole proprietors, partnerships, and subchapter S-corporations, asking for information about business owner demographic characteristics, the level and source of start-up financing, and the importance of the business as a source of income to the owner. This is valuable information for understanding the business formation process, but the data are collected infrequently. To fill this gap, the Federal Reserve sponsors the Survey of Small Business Finance because existing data on bank lending and financing to small businesses are inadequate. The survey is quite small and voluntary in nature, so it has both response and coverage issues—we also understand there are plans to discontinue it which, despite these shortcomings, would be a significant loss.

4.3 SYSTEMIC DEFICIENCIES

4.3.1 Insufficient Interagency Data Coordination

In some cases, the data that would be most useful to researchers and policy makers for measuring business dynamics are simply not collected. In other cases, survey data are collected or administrative records maintained, but they cannot be shared among the statistical agencies.[13] Insufficient data

[12]The SBO is similar to the Characteristics of Business Owners survey conducted in prior census years (e.g., 1987 and 1992—the survey was cancelled in 1997 due to budget limitations). A number of rich studies have been conducted using Characteristics of Business Owners and SBO data sets—see, for example, Holmes and Schmitz (1995).

[13]See Appendix 4-1 for a brief history of interagency data sharing.

sharing and lack of coordination result in an underutilization of resources. Not only can data sharing result in improved quality of economic (and other) statistics, it may also potentially reduce costs to the agencies and the burden placed on survey respondents who receive similar data requests from multiple agencies (see Chapter 3 on this point).

The production of economic statistics relies heavily on survey and administrative data collected and housed by the Census Bureau, BLS, and BEA. However, with a few specific exceptions, these agencies are not permitted to share their data sources when they include Title 26 (U.S. Code) federal tax information. The Confidentiality Information Protection and Statistical Efficiency Act (CIPSEA) has expanded the potential for data sharing among these agencies. However, enabling legislation for the sharing of tax data is not part of CIPSEA. As a result, data in the Census Bureau's business register, which is constructed in large part from IRS tax data, cannot be shared.[14] This, in turn, impedes coordination of the business lists maintained by the statistical agencies. In some cases, agencies have had to purchase private business lists that are inferior in some ways, simply because they did not have access to the business data at another agency (for example, the Survey of Small Business Finances at the Federal Reserve has used sampling frames from Dun & Bradstreet). Controlled access to government lists would eliminate these kinds of costs and allow more accurate data to be used.

Since the Census Bureau's business register relies heavily on IRS source data, the agency's ability to share with BEA and BLS is extremely limited. The Census Bureau itself does not collect any data on receipts from nonemployer businesses but simply uses IRS data for such cases. These data cannot be shared. Likewise, BLS and BEA access to data on sole proprietors—of which there are roughly 1.5 million, a group of businesses that constitutes a large fraction of economic activity—and partnerships is extremely limited and has been excluded from previous data-sharing proposals.[15]

---

[14]See Appendix B in the Committee on National Statistics (CNSTAT) report *Improving Business Statistics Through Interagency Data Sharing* (National Research Council, 2006) for an overview of recent legislation governing data sharing and access to federal tax data. Papers by Nick Greenia and Mark Mazur and by Robert Parker in that volume provide further details of statistical agency access to IRS data. The full text of the CIPSEA legislation can be found at http://www.eia.doe.gov/oss/CIPSEA.pdf.

[15]The effort to promote greater data sharing for business list reconciliation is further complicated because BLS data are shared with state programs. The Labor Market Information Cooperative Agreement includes provisions for BLS and state agencies to share data for five BLS programs—Current Employment Statistics, Local Area Unemployment Statistics, Occupational Employment Statistics, the QCEW, and Mass Layoff Statistics. Under the legal terms of the agreement, the state agencies have access to data collected through these

Inability to fully share business register microdata creates problems for a number of statistical programs and research needs. As noted above, it harms the accuracy of industry output, compensation, and productivity trend measures (especially for fast-growing and innovative industries, such as information technology) which, in turn, affects the ability of downstream users, such as the Federal Reserve, BEA, and researchers, to use the data effectively. It also inhibits the ability of the statistical agencies to consistently adjust sampling frames for the entry and exit of new businesses in a timely manner and to identify and classify businesses (again, see National Research Council, 2006). For example, both the Census Bureau and BLS require establishment-level data for multiunit firms. The Census Bureau requests multiestablishment firms to break out employment and payroll numbers by establishment in its COS; for BLS, the UI program's MWR is used (refer to Appendix A for details). Because the timeliness and comprehensiveness of the COS and MWR are not the same, measures of employment, payroll, and establishment birth and death trends for multiunit firms are incomplete and inconsistent.

While the statistical agencies already share data—effectively in some cases—a more extensive data-sharing arrangement between BLS and Census would likely lead to improvements in both lists. BLS industry coding, physical location addresses, multiunit data from the MWR, and employment data for single units are recognized as being very thorough, and this level of detail would be—and, to some extent, already is—beneficial for use in Census programs.[16] The Census Bureau is particularly interested in the data of multiunit companies within states, as well as in BLS data for the client businesses of professional employer organizations (PEOs). PEOs (or employee leasing) firms typically supply human resource management services (e.g., payroll accounting, benefits administration) to their clients. The Census Bureau's tax record-based data do not accurately indicate the geographic location and industry of leased employees working at client sites; rather, they indicate the industry and location of the PEO itself. BLS would benefit from an evaluation of firm information that is collected as part of the Census Bureau's COS. Access to Census data could potentially add

programs, and they agree to use confidential information for statistical purposes. Authorized state employees can share microdata within the agency, but they are under the same limitations as BLS employees in terms of what can be publicly released (Bureau of Labor Statistics, 2006).

[16]The Census Bureau receives over 3 million industry codes, county codes, employment figures, and physical location addresses each year from BLS. While this data-sharing project reduces costs for the Census Bureau and relieves respondent burden, much of the potential benefit to the statistical system is left untapped since BLS cannot obtain IRS-sourced data from the Census Bureau.

consistency to BLS industry codes, giving the agency the ability to analyze microdata on some 18.6 million nonemployer businesses. A workshop held on the topic (see National Research Council, 2006) explored possible mechanisms for increasing sharing for the purpose of business list improvement; Chapter 5 includes recommendations on how to move forward.

Integration of the business registers and the business surveys is a complex and difficult process. Sorting through what matters for key national statistics and for understanding the driving forces of the U.S. economy is a long-standing topic of mutual interest to the statistical, research, and policymaking communities. Improving measurement and understanding of business dynamics in general, and young and small businesses in particular, can be accomplished only through partnerships between the statistical agencies and the research community. This in turn requires access, with appropriate safeguards and for approved projects, to the underlying firm-level data. This partnership is an essential ingredient to tackling conceptual and measurement problems because the statistical agencies have limited resources to address these issues by themselves.

4.3.2 Inadequate Researcher Access

Even the best data are without value if no one can use them. Another limitation of the current system has to do with who can use data beyond the walls of government agencies. Our primary concern is with research and policy users, but data access by others, including business planners, is also important. Because the vast share of expertise in data analysis resides outside the statistical agencies (which are, for the most part, working near capacity just to produce the data), academic researchers and government policy makers must be given access if the data are to be used with an intensity sufficient to justify the costs of collection. The importance of researcher access to data, as well as the confidentiality requirements that must be met in order to provide that access, has received extensive attention in several CNSTAT (and other National Research Council) reports. A major theme of *Expanding Access to Research Data* (National Research Council, 2005b) is that a full return on the nation's investment in the federal statistical system requires that users have access to data which, in turn, informs policy and feeds back to improve the quality of surveys and of the data sets (public use and restricted access) themselves (pp. 48-49):

> Researchers' use of government data creates an effective feedback loop by revealing data quality and processing problems, as well as new data needs, which can spur statistical agencies to improve their operations and make their data more relevant. . . . McGuckin (1992) argued more than a de-

> cade ago that coordinated research efforts between in-house and outside researchers offer the best model for ensuring that agencies maximize the benefits from data users. In fact, McGuckin (1992:19) argued that it is a primary responsibility of statistical agencies to facilitate researcher access to confidential microdata files. Such access, by improving the microdata for research and policy analysis, will also improve the quality and usefulness of the aggregate statistics on trends and distributions that are the bread and butter of statistical agency output.

Researcher access is critical for addressing the data gaps highlighted in this chapter. Data sharing with accompanying data integration and coordination within the statistical agencies will go a long way toward improvements in business statistics. However, many difficult conceptual and measurement issues must be confronted, and a partnership between the statistical agencies and the research community is vital to such efforts. It is useful to note that major innovations in measuring business dynamics have stemmed from partnerships. The methodology underlying the job creation and destruction measures now part of the BED and the LEHD programs was developed under joint projects between Census Bureau staff and external researchers at the Center for Economic Studies (see Davis, Haltiwanger, and Schuh, 1996, for discussion of this partnership). A joint project between John Abowd and the French National Institute for Statistics and Economic Studies (see Abowd, Kramarz, and Margolis, 1999, and Abowd, Corbel, and Kramarz, 1999) and a related joint project between Julia Lane and the Census Bureau (see Burgess, Lane, and Stevens, 2001, and Haltiwanger, Lane, and Spletzer, 1999) led to critical conceptual and measurement breakthroughs that underlie many of the key data products of the LEHD program.[17]

Throughout this report, we document the value of business data, particularly when it can be disaggregated along geographic and other dimensions. However, as noted in Chapter 3, the presence of geographically specific identifiers along with other variables in a database gives rise to confidentiality issues.[18] Social scientists have begun to grapple with the

[17]It is also worth emphasizing that the joint statistical agency and research community projects discussed here have been jointly funded by the statistical agencies and major public and private organizations, including the National Science Foundation, the National Institute on Aging, the Sloan Foundation, and the Kauffman Foundation.

[18]A 2001 article reporting the results of a survey conducted by the Urban Institute indicates that business taxpayers are willing to allow more access to statistical agencies for some types of tax return information (Greenia, Lane, and Jensen, 2001). Of particular interest to the issue of privacy interests is the response to a set of questions about types of data businesses view as very sensitive. The survey results showed that less than 5 percent of respon-

dilemma posed by the conflicting needs for spatially explicit data and confidentiality (Van Wey et al., 2005).

A number of strategies for meeting the dual goals of confidentiality protection and user access are possible (Box 4-2 includes a description of an approach used by Statistics Denmark):

- Sensitive disaggregated data can be held by a central agency and access to sensitive databases restricted to researchers and policy makers who have cleared security hurdles.
- Disaggregate data can be held by a central agency that responds to user requests for data for certain areas and releases the data only in aggregated form. Because confidentiality can be breached if the data are released for small areas such as census blocks, block groups, or even census tracts, the minimum size of area for data release remains an empirical question that can be answered only in terms of the probability that a given business can be identified, which is a function of the density of similar businesses in an area, inter alia.
- The location identifier in the disaggregate database can be generalized from a point to an area; in this strategy, location specificity is sacrificed by "geo-fuzzing" to preserve confidentiality. As in the second case, the size of the area to which the location needs to be generalized to maintain confidentiality depends on the spatial distribution of proximate similar businesses (compare, for example, a business in a rural area—the only one of its kind in a county—with a similar one in a dense urban area that has many similar businesses within the same census tract).

The problem of balancing increased access to microdata while conforming to confidentiality requirements has also led to the idea of creating synthetic versions of microdata sets. The development of methods for gen-

---

dents thought their name and address and industrial activity were very sensitive, but 85 percent of respondents were very sensitive to data about their employees. From 50 to 75 percent of the respondents thought their financial data were very sensitive. Smaller firms seem to indicate that financial information is more sensitive than do larger firms. Smaller firms are consequently more concerned than larger firms about providing information to the Census Bureau and other statistical agencies; multiunit companies consider company-level data more sensitive than establishment-level data. Willimack, Nichols, and Sudman (2002) note that "businesses that are more dependent on their [external] environments, such as publicly-traded firms, have higher motivation to disclose information, while those that are insulated or in unregulated environments are more protective of information. . . . Those in more volatile industries [or markets] were more protective of their data, because releasing information could result in a loss of competitive advantage" (p. 223).

BOX 4-2
Data Access Abroad

Data access arrangements differ significantly from country to country. In some cases, recognized researchers can obtain data samples on their own computers (to some extent, this is the case in Norway and Sweden). In others cases, researchers must travel to statistical agencies to gain access (Germany); this is of course highly inconvenient and, by lowering use rates, reduces the societal value of the data resource.

Denmark has a special arrangement worth mentioning. The Ministry of Research together with Statistics Denmark set up a system whereby researchers can remotely access data sets they have ordered through the agency. Researchers apply to use data for a specific project, then Statistics Denmark creates a data set specified in a way that meets the criteria for that use. In constructing the data set, the agency may allow use of the full range of available data. However, it is up to their discretion to determine the extent to which a researcher's information request is fulfilled. The final step in data delivery involves anonymizing the data then moving them to the researcher's computer system. Researchers are of course not allowed to publish or to take out any data that can be used to identify individuals or firms. The arrangement does allow them to gain remote access to data from their offices over a virtual network connection. Local printing is not possible, nor is downloading of data. Output can instead be sent as email. These emails are screened by Statistics Denmark employees or by a computer program for any breach of data security before they are sent, which means a delay of about 5 minutes. The system has been working for about 5 years without incident. One downside of this arrangement is that most researchers use seemingly different data sets, with the result that typical scientific critique can be more difficult and discussion of data limited to some extent.

erating synthetic versions of data sets is relatively recent, dating back to Rubin (1993). The synthetic data approach is essentially data masking, though it is more sophisticated than such conventional techniques as rounding, top-coding, withholding variables, and creating value ranges.[19] Methods of disclosure limitation based on synthetic or virtual data have shown promise in safeguarding confidentiality and permitting the estimation of

[19]The idea of synthetic samples is that one has available a sample of size n with observations on X and Y, and a much bigger sample with N observations on X alone. This may be because data on Y are confidential. The small sample can then be used to construct a model for the conditional distribution of Y given X. Draws from this distribution are then made for each observation on X in the larger sample. The result may be useful if the relation between X and Y is tight and the distribution of X in the larger sample is quite different from that in the small sample. Even if the model is not a particularly tight fit, the data set may be useful if it is sufficiently broadly conditioned on the observed data and if the standard errors can be adjusted to reflect the data synthesizing step.

complex models (see, e.g., Abowd and Woodcock, 2001; Doyle et al., 2001; Raghunathan, Reiter, and Rubin, 2003).

From a research perspective, synthetic data sets are not ideal; rather, they are a compromise. There have always been two perspectives when it comes to synthetic data: on one side are those who believe that, in addition to its role in statistical disclosure limitation, replacing real samples with records created from posterior distributions offers significant potential in terms of maintaining the analytic value of the data. On the other side are researchers who are concerned that imputation algorithms blur data in ways that introduce unacceptable effects on error structures (and other inaccuracies), particularly in complex modeling applications. In general, social scientists have concluded that the extent to which a synthetic sample representing "virtual households" or "virtual businesses" is useful depends on the particular implementation of the method and on the particular questions being addressed.

However, it is extremely important to note that the method offers a highly valuable data access tool. Synthetic public-use microdata versions can be used to facilitate access to non- (or less) sensitive versions of restricted access data sets. This kind of access can provide a means for researchers to explore, test, and refine estimation models at relatively low cost before incurring the higher costs of accessing confidential data through a research data center or another restricted access mode. Synthetic data and multiple imputation techniques are relevant to several of the data sets discussed in this report. One example is the multiple imputation apparatus in place for the Survey of Consumer Finances (SCF). The SCF is conducted by the Federal Reserve Board, with survey information collected by the National Opinion Research Center at the University of Chicago. The sampling frame is constructed from records based on tax returns maintained by the Statistics and Income Division of the IRS. Because the SCF contains sensitive and detailed information about respondents' assets and liabilities, as well as detailed demographic and geographic codes, and because missing data have always been an important problem in the SCF, substantial resources have been devoted to the construction of an imputation framework that can be used to simulate data to replace those originally reported.

Synthetic data and multiple imputation techniques have also been used in the LEHD program. This has allowed data to be released at the block level without compromising the confidentiality of individuals or businesses. One application of this method, currently being used in 14 states for economic and workforce development, permits decision makers to examine job creation, job destruction, hires, and separations at the block group level—as well as their dynamics over time (see http://lehd.excensusonline.com).

The use of multiple imputation has been helpful to address the problem

that UI wage records typically do not include the physical establishment for workers employed by multiunit businesses. For these workers, only the identifier for the multiunit business is recorded. The LEHD program staff developed a multiple imputation technique to assign a place of work to these individuals based on the size and hiring patterns of establishments within the multiunit business and the relationship between the place of residence of each worker and the location of each business. This is an example of using imputation to improve the quality of data. Imputation can therefore be useful for two purposes—when there are missing data and when confidentiality must be protected.

4.4 APPENDIX: DATA-SHARING HISTORY[20]

The potential benefits of more efficient data sharing among the U.S. statistical agencies have long been recognized. The National Research Council's 1993 report, *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*, included a section on "Barriers to Data Sharing within Government" in which the authoring panel wrote (p. 6):

> Some of the laws that govern the confidentiality of statistical data prohibit or severely limit interagency sharing of data for statistical purposes. Laws that control access to administrative records, such as reports of earnings covered by Social Security, restrict their use for statistical purposes. These barriers to data sharing for statistical purposes have led to costly duplication of effort and excessive burden on individuals and organizations who are asked to supply information. They have also made it difficult or impossible to develop data sets needed for policy analysis on topics of major interest to the public.

The panel concluded that barriers to sharing data on both persons and businesses for statistical purposes should be removed subject to strict controls to protect confidentiality. Recommendation 7.4 of that report articulates the panel's position that interagency data coordination should include the sharing of lists of businesses by federal and state agencies for statistical purposes. Recommendation 7.5 asked for new legislation to expand confidentiality to records collected by all statistical agencies; some of this recommendation was subsequently accomplished by CIPSEA. Although this report cited several examples of reduced sharing as a result of the provisions

[20]Much of the factual background on data-sharing legislation in this section is reproduced from an article written by Robert Parker for a CNSTAT report (National Research Council, 2006).

of Title 26 (specifically, Section 6103) of the tax code, there was no explicit recommendation to amend this act. Nevertheless, the panel's view on the topic was clearly expressed by a section titled "Inability to Share Business Lists: An Embarrassment to the Federal Statistical System."

A more recent CNSTAT report, *Expanding Access to Research Data: Reconciling Risks and Opportunities* (National Research Council, 2005b), includes a detailed discussion of CIPSEA and specifically points to the need for changing Section 6103 (p. 23):

> A key element in the Census Bureau's data is its business register, which is constructed with data from the Internal Revenue Service (IRS). However, without new legislation (to amend Title 26 of the U.S. Code, which governs access to IRS tax data), the business register and associated data cannot be shared with BEA and BLS.

Many organizations have weighed in with positions supporting amendment of Section 6103 of the tax code. In response to a congressional request, the Administrative Conference commissioned a team of tax experts headed by Charles Davenport to study the operations of the IRS. The result was the *Report on Administrative Procedures of the Internal Revenue Service, October 1975, to the Administrative Conference of the United States* (Davenport, 1976), which included a section reviewing the history and rationale for tax return access for statistical purposes. The report concluded that "it appears that the use of tax data by Census, though not consistent with revenue administration, is a use which can be considered beneficial and is one which does not appear to have any undesirable side effects" (p. 880). The study came to a similar conclusion for BEA but not for the Statistical Research Service (SRS) of the Department of Agriculture. The key distinguishing factor for the commission was that the Census Bureau and BEA were strictly statistical agencies and did not engage in activities related to other functions of the department. Still, the commission recognized that SRS had statistical activities but determined that the statistics it collected "for policy making by the agency of which it is a part."

In 1979, the U.S. General Accounting Office (GAO, now the Government Accountability Office) issued a report identifying efforts by the Census Bureau to create a centralized business register for use by other statistical agencies. That report, *After Six Years, Legal Obstacles Continue to Restrict Government Use of the Standard Statistical Establish List*, supported changing Section 6103; its findings were summarized as follows (title page):

> The Census Bureau has developed the Standard Statistical Establishment List, a comprehensive list of businesses in the United States. Many Federal

> agencies could use such information. But confidentiality laws prevent the Census Bureau from sharing List information with other agencies. . . . Amendments to these laws would help improve the quality and comparability of economic statistics and reduce business response burden from numerous Federal statistical surveys. . . . Because some of the List data comes from the Internal Revenue Service, the Treasury Department has reservations about using tax information for statistical purposes. However, the Commerce Department plans to introduce proposals for changes to the confidentiality laws and GAO recommends favorable congressional consideration.

In 1998, GAO testified before Congress (based on its report, *Statistical Agencies: Proposed Consolidation and Data Sharing Legislation*) about legislation submitted in 1996 and 1997 that would permit limited sharing of data among designated statistical agencies for statistical purposes, subject to procedural safeguards. The testimony included the following statement describing GAO's position (pp. 5-6):

> For the past 2 decades, we and others have urged legislative changes that would allow greater sharing of data and information on data sources among agencies, but so far these efforts have met with little success. The Paperwork Reduction Act of 1980 gave the Director of OMB the authority to direct a statistical agency to share information it had collected with another statistical agency. However, this authority was limited since it did not apply to information that was covered by laws prohibiting disclosure outside the collecting agency. In the early 1980s, the statistical agencies, under OMB's leadership, tried to further enable federal statistical agencies to share data. They attempted to synthesize, in a single bill, a set of confidentiality policies that could be applied consistently to all federal agencies or their components that collected data for statistical purposes. This effort, which was known as the "statistical enclave" bill, would have allowed statistical agencies to exchange information under specific controls intended to preserve the confidentiality of the data providers. A bill was introduced in Congress but was not enacted.
>
> More recent proposals concerning data sharing have called for enactment of legislation that would allow statistical agencies to share data and information with appropriate safeguards to protect against breaches of confidentiality. These proposals were not adopted, in part because of general concerns that greater data sharing might endanger the privacy of individuals. Both the Economic Statistics Initiative under President Bush and the National Performance Review (NPR) under President Clinton have recommended such actions. NPR recommended the elimination of legislative barriers to the exchange of business data among federal statistical agencies, and we agreed with this recommendation. The NPR recommendation did not address the sharing of information on individuals.

> Some officials of statistical agencies and Members of Congress, however, have argued that a distinction should be made between the sharing of business data and the sharing of personal data about individuals. They noted that breaches of confidentiality protection when personal information is involved may be more serious. The National Academy of Sciences has made recommendations regarding the need for appropriate legislative provisions on data sharing that the Subcommittee may wish to consider in its deliberations.

In the 2002 *Economic Report of the President*, the Council of Economic Advisors noted the critical need for reliable data and that data sharing would increase their quality and timeliness (p. 25):

> . . . the quality of existing statistics is far from perfect and could be enhanced with further investment. Even real GDP, generally thought of as a reliable measure of overall activity in the U.S. economy, is susceptible to considerable revisions. . . . Such revisions lead to uncertainty for both government and private decision makers, which can cause costly delays. . . . A number of steps can be taken to improve the accuracy and timeliness of economic statistics. In particular, targeted improvements to the source data for the national accounts would go a long way toward illuminating the causes of the growing statistical discrepancy. Another cost-effective measure would be to ease the current restrictions on the sharing of confidential statistical data among federal statistical agencies. Such data sharing, which would be done solely for statistical purposes, is currently hindered by lack of a uniform confidentiality policy. Confidentiality is of key importance to all agencies and to the individuals and businesses who participate in Federal surveys, but a uniform confidentiality policy would allow agencies such as the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Census Bureau to cost-effectively compare and improve the quality of their published statistics while preserving confidentiality. In the past, attempts have been made to pass legislation, together with a conforming bill to modify the Internal Revenue Code, allowing such data sharing under carefully crafted agreements between or among statistical agencies. In 1999 such legislation passed the House but stalled in the Senate. The Administration will continue to seek passage of data sharing legislation to improve the quality and effectiveness of Federal statistical programs.

Positions on data sharing have also been issued by the Treasury Department and the Joint Committee on Taxation (JCT). Section 3802 of the Internal Revenue Service Restructuring and Reform Act of 1998 requires the secretary of the treasury and the JCT to conduct separate studies of the scope and use of provisions regarding taxpayer confidentiality and to report their findings, together with any recommendations deemed appropri-

ate, to Congress. The JCT published its report on January 28, 2000; the Office of Tax Policy of the Department of the Treasury submitted its report on October 2, 2000.

The JCT and the Treasury Department have disagreed on the merits of changing Section 6103 to permit expanded data sharing. The JCT recommended that "new access to returns and return information should not be provided unless the requesting agency can establish a compelling need for the disclosure that clearly outweighs the privacy interests of the taxpayer." The JCT report did not explain how to determine both the pros and cons of such a decision, nor did it appear to fully accept the need for the existing access. (Both the JCT and the Treasury Department recommended dropping the Federal Trade Commission from access under Section 6103.)

The Treasury Department, in contrast, recommended that "the disclosure authority of Section 6103(j) should be expanded to additional specified statistical agencies, and such agencies should be permitted, upon prior Treasury approval, to share IRS data with each other." It also specified the agencies to include and that the change to Section 6103 should cover both individual and business records.

Private business economists have actively supported data sharing. In 1996, Maurine Haver, president of Haver Analytics and chair of the Statistics Committee of the National Association for Business Economics (NABE), testified before the House Subcommittee on Government Management, Information and Technology. In her testimony on consolidating the three major economic statistical agencies (H.R. 2521), she expressed support for the inclusion in that bill of provisions for data sharing among these agencies as it would allow for the creation of a single business register. In 2001, Richard Berner, the president of NABE and managing director and chief U.S. economist of Morgan Stanley, Inc., testified before the House Subcommittee of the Census Committee on Government Reform on the Statistical Efficiency Act of 1999. In his testimony, Berner reported on NABE's support for the reintroduction of this act, which had been passed unanimously by the House in the previous Congress. He testified that "NABE believes that our national data collection efforts should be as efficient as possible. To that end, we believe that Congress should mandate 'data sharing' among the agencies, solely for statistical purposes." He noted that existing confidentiality statutes are barriers to such data sharing because "they virtually guarantee duplication of effort and inconsistencies among related data sets collected by the affected agencies. Moreover, they prevent agencies from undertaking new analyses that could improve the information available to policy makers. This is not a cost-effective way to run any business—either public or private."

5

Improving Data and Statistics on Business Dynamics—Bridging the Gap Between the Current and a Comprehensive System

In this report we argue that the constellation of business data currently produced by the statistical agencies, while impressive in many respects, could be substantially improved. An increased focus by statistical programs on business dynamics would facilitate a more complete understanding of business creation processes, of the mechanisms whereby firms adapt and change with the economy, of the role of new and young firms in economic growth, of how new sectors emerge and new markets are created, and of shifts in employment opportunities across sectoral and geographic dimensions. Additionally, while the United States is widely considered the exemplar of an "entrepreneurial nation"—largely because of its role in the creation of entirely new markets and economic sectors—there is very little precise information about how this mechanism actually works and what policy actions might enhance (or harm) this environment.

In this chapter, we provide a framework and recommendations for (1) improving the representativeness and quality of a broad range of business surveys, (2) generating more timely descriptions of changes in the U.S. economy that allow policy makers to respond more quickly to the changing business environment, and (3) expanding the scope and details of information on individuals businesses. We offer specific recommendations—about how to improve the business lists residing at the statistical agencies as well as other data sources relevant to the measurement of business formation, dynamics, and performance—while recognizing the need to minimize additional costs and respondent burden. In constructing our recommendations,

we attempt to differentiate between strategies that could be implemented relatively quickly at modest cost and those that would require longer term commitment.

Early on, the panel identified a set of principles to guide its work and, in turn, the development of recommendations presented in this chapter. These principles are detailed in Chapter 1 and summarized here:

- *Confidentiality*: Statistical agencies have the responsibility to data providers and data subjects to protect the confidentiality of information that is provided. Data collected by the government must be maintained in such a way that identifiable information is not disclosed for administrative, regulatory, or enforcement purposes.
- *Public Purpose*: Subject to confidentiality requirements, data sharing among government statistical agencies and data access by others should be facilitated when it serves a substantial public purpose. Data uses that serve a substantial public purpose include those that (1) lead to improvements in the quality, breadth, and usefulness of government statistical data and systems; (2) provide evidence crucial to informing government policies on social and economic issues; and (3) encourage research that advances scientific knowledge. The rationale for the public purpose principle is straightforward: government administrative record systems and survey databases generate enormous public value in terms of informing decision makers (including those in the private sector) and are maintained at considerable cost to the public. As such, the public is entitled to the full and effective use of these assets, provided that such uses do not compromise the confidentiality assurances afforded to respondents.
- *Targeting Deficiencies*: Improvements to data collection should focus first on areas for which policy and research relevance is high but statistics needed to inform those policies and research are weakest. For business data, this means building up the statistical infrastructure for measuring dynamics and collecting information on rapidly growing economic sectors in which the activities of smaller and younger firms are disproportionately important, but for which data coverage is relatively weak.
- *Cost Efficiency*: The statistical agencies should give the highest priority to actions that can be done expeditiously and at low cost. In this report, we identify a number of cases for which more creative use of existing data could lead to the production of useful statistics. The idea is to get as much information out of the system as possible for a given level of resource and data protection commitment.

Reflecting the charge to the panel and the concentration of its efforts, recommendations are organized around three systemic needs:

- increasing the capacity to measure activities of nascent and young businesses that rapidly enter and exit fast-growing and innovative sectors of the economy and that are central to understanding business dynamics;
- improving the coverage and depth of business data through more effective coordination and integration of existing sources; and
- shifting the legal and organizational environment to accommodate data sharing and confidentiality protections in a way that enables the kinds of efficiencies envisioned by the panel to occur.

5.1 EXPANDING DATA SOURCES FOR MEASURING BUSINESS DYNAMICS

Although accounting for only a small portion of economy-wide revenues, nonemployers (most are sole proprietors) and other small firms represent the vast majority of businesses in the United States. More importantly, these microbusinesses appear to disproportionately contribute to changes in the composition of the economy's product and labor markets. A small percentage but large absolute number of these businesses evolve into firms with employees. Data on nonemployers, sole proprietors, and those involved in entrepreneurial activities are therefore essential for studying business dynamics.

Measuring key business and worker transitions typically requires longitudinal data covering the early and late life phases of businesses. Ideally, entities are tracked at the establishment level in such a way that changes (for example, in the kinds of goods or services produced) can be detected, even when location and name remain the same. In addition, data are needed that can be disaggregated to local levels. Recent examples of high-profile, localized economic transitions include the workforce mobilization out of New Orleans following Hurricane Katrina and those associated with military base closings and realignments. High-quality samples that are representative of a broad range of businesses—old and new, large and small—would allow for more timely analysis of these kinds of events and, in turn, provide opportunities for policy makers to respond more quickly to changing market conditions. An ideal business data system would facilitate measurement of the attributes that are linked to (and possibly predictive of) business performance, outcomes for individuals and communities, and local economic trends.

5.1.1 Sampling Young and Small Firms

In designing a data collection system, nothing is more fundamental than the question of whom to survey. The optimal mix of business entities

to be covered in the statistical system's surveys, censuses, and administrative sources must be determined. The challenge is that this is a multipurpose problem—there is a trade-off between measuring levels and measuring dynamics. Because it is important to estimate the volume of output, employment, and other variables at low cost, statistical programs have historically focused on the largest entities. However, to accurately measure changes occurring among businesses and in markets, which for many purposes is more important, data must also be collected on emerging and fading entities—typically the smallest.

For measuring business dynamics, it would be beneficial to reduce the undersampling of those parts of the business population that are most likely to be in transition and that provide early indicators of the future directions of the economy.

Recommendation 1: To measure business dynamics more effectively, the Census Bureau and the Bureau of Labor Statistics (BLS) should increase the sampling of younger units in their surveys. This will require that business age is included as one of the stratifying variables and that business lists, on which the surveys are based, cover recent business entrants.

When optimizing a sampling structure, there are trade-offs in terms of picking up changes in variables versus achieving precision in population total estimates. Currently, most survey programs stratify by industry and size, and samples are chosen to minimize the sampling error of level estimates. This approach leads to lower (in most cases, much lower) sampling probability for small businesses relative to large ones. If the statistical agencies move to instead minimize a criteria function that includes sampling errors for both levels and growth rates—for example, some weighted average of the mean squared error—then this would increase the sampling probability of small businesses. Stratifying by age, as we suggest, will have similar effects. Although the case for these recommended changes is strong, implementing them would require a rethinking of the fundamental approach by the agencies to business sampling. Acting on this recommendation would entail new costs and take time to implement but, with very little new expenditure, the agencies could immediately begin undertaking research to quantify the statistical trade-offs associated with adjusting sampling rates of businesses along age and size dimensions.

As things now stand, nonemployer firms—many of whom are initiated as sole proprietorships or partnerships—are almost completely unrepresented in federal data programs, yet these businesses are frequently associated with the most dynamic elements of the economy. That sources of microdata have historically been scarce for nonemployers, and for smaller and newer businesses more generally, has hampered research progress on

business dynamics. Recently, however, data products have emerged that promise to greatly enhance available information relevant to the topic. Because of their proximity to the business lists, the Census Bureau and BLS are the key architects in this emerging data area; their efforts should be applauded and further development of these sources encouraged.

Recommendation 2: BLS and the Census Bureau should support and expand their development of statistical programs such as the Business Employment Dynamics (BED) and Statistics of U.S. Businesses (SUSB) that provide basic measures of business dynamics, including statistics on business formation and dissolution and job creation and destruction.

Some extensions of these programs would admittedly necessitate longer term commitment, while others may be initiated by more intensive use of existing data. For example, the Census Bureau could add significant value to the SUSB by incorporating information on the dynamics of nonemployer firms (taking advantage of the development of the Integrated Longitudinal Business Database or ILBD). This would clearly be an expansion, albeit a useful one, of the program. Still, progress could be made on some fronts with relatively little additional cost. BLS could accelerate the development of more disaggregate tabulations at the geographic (substate) and industry (6-digit NAICS) levels with little or no new data collection, just more aggressive processing of administrative records data.

If one accepts the premise that data on small and young businesses are inadequate and that important research and policies rely on such information, then it follows that key data programs must keep careful tabs on how long business entities have existed.

Recommendation 3: The Census Bureau and BLS should exploit their administrative record systems to produce public-release statistics on economic activity disaggregated by indicators of business age. Readily available business age indicators in these administrative records systems include the application date for an Employer Identification Number (EIN), the point at which positive revenues are generated, and the first period with positive payroll.

Acting on this recommendation will require only modest adjustments to existing data collection instruments; indeed, age tables were produced for the Census of Retail in 1939 and 1948. The Census Bureau and BLS currently publish numerous public-release statistics disaggregated along other dimensions—for example, data on productivity by industry and firm size. It would be similarly useful, in terms of monitoring comparative trends of new entrants in the economy, if statistics were maintained for firms and

industries by age. The appropriate milestone for defining a business birth will vary by purpose.

A focus on publishing statistics by business age would also dovetail well with the recent innovations in measuring producer dynamics in the microdata-based BED and SUSB/LBD programs. Since these programs rely on the accurate measurement of entry and exit of producers and the accurate tracking of existing producers over time, their statistical frames could be readily adapted to include statistics disaggregated by business age.

The development of longitudinal versions of the business registers at both the Census Bureau and BLS would permit using business age as a stratifying variable to annual, monthly, and quarterly surveys. Because many key statistics (e.g., productivity by industry) integrate survey information from multiple sources, adding downstream data products delineated by business age would require increased coordination by the agencies to make definitions consistent. Initially, business age should be added to surveys for which the new information would be most valuable. Good candidates for this might be the Annual Capital Expenditures Survey and the National Science Foundation's Research and Development Survey. The Census Bureau's Survey of Business Owners (SBO) offers something of a model, given that it already asks respondents for information on business age.

5.1.2 Nascent Business Activity—The Essential Role of Household-Based Data

Tracing the entire life cycle of businesses and measuring and analyzing the processes through which they are born and grow require going beyond traditional data collection from employer businesses. Nascent businesses encompass the entrepreneurial activities of individuals or households before they come in contact with the legal system as business entities—thus, business registers take one only so far in measuring business dynamics. Only after acquiring an EIN as a federal business taxpayer, or as a state Unemployment Insurance taxpayer, is a business tracked in the frames used by BLS and the Census Bureau to measure economic activity. Unlike the BLS register, the Census Bureau register does includes nonemployer businesses if they have taxable revenues. However, businesses are not typically included in the major surveys of these agencies until they become employer businesses with positive payrolls or taxable revenues.

There is little in the current system that provides a way of tracking individuals as they enter into the business creation process and spend time and resources in an effort to organize and implement a new firm. The most direct way to capture many of the activities (and characteristics of those

carrying out the activities) associated with the early stages of business formation is to collect data from household or individual units.

While there are limitations to household-based data, such as the typical absence of information on business performance, they can add unique analytic capacity for understanding business dynamics. Ideally, information collected from households would be linkable to business data sets through unique identification numbers. Integrating data from households and from employers is critical—perhaps increasingly so—for tracking the growth of new firms and emerging sectors, and for developing a more complete picture of employment flows in the economy.

Existing household surveys could be used as the screening vehicle for identifying nascent and young businesses.

Recommendation 4: The Census Bureau should periodically add a module to the American Community Survey (or possibly the Current Population Survey) to identify nascent entrepreneurs. A method should be developed for linking this survey information with subsequent business identifiers in a longitudinal household-business data infrastructure so that transitions from nascent to active status (and vice versa) and from nonemployer to employer status (and vice versa) can be measured and studied.

Adding a module or a screener question to household surveys can, in principle, capture activity associated with nascent businesses. Such a plan must acknowledge that the majority of households will not be populated with individuals involved in business start-ups. Because the sampling frame is not particularly efficient for locating such activity, only very large surveys would generate sufficient numbers of eligible cases. That said, only minor modifications to the structure of items asked in the Current Population Survey (CPS) or the American Community Survey (ACS) would be required. These modifications could be implemented immediately with relatively little additional monetary cost, though we recognize that there is an opportunity cost (given the widespread interest in adding content to household surveys, there are a number of topical modules worthy of consideration).

The Census Bureau should also consider implementing a program of periodic follow-on surveys of the screened nascent entrepreneurs and young businesses using specialized topic modules. Combining responses to a well-designed set of questions with some longitudinal follow-ups and administrative record linkage would provide a pathway for studying the dynamics of these businesses over their entire life cycles.[1] If these individuals could be

[1]Approximately one-third of new initiatives ultimately become incorporated into the various business registries.

followed over a longer period of time—up to three or four years—activity could be tracked until their businesses entered the system through a Schedule C tax return filing or, if it had employees, via the Unemployment Insurance system records. Such modules could be rotated over time to cover a range of both firm and business owner variables. Data from these modules would provide estimates of the prevalence of independent start-ups and business sponsored start-ups among the adult population and should be stored in a database that facilitates longitudinal analyses.

5.1.3 Surveying Business Owners

In addition to tracking changes that business entities undergo, it would also be beneficial to be able to monitor transitions that accompany the earliest phases in the lives of the owners who start them. Given the focus of many surveys on large producers, timely information on start-up financing, human resources, and investments in research and developoment and physical capital is often inadequate for young and small firms. This is particularly true for the nonemployer segment of the business population. One survey vehicle that does provide coverage of both the employer and nonemployer universes is the SBO. A key feature of this survey is that it identifies business age. The SBO generates statistics on the composition of U.S. businesses and on owner characteristics. Economic policy makers in federal, state, and local governments use SBO data as a source of information on business success and failure rates. The survey is particularly useful for comparing the performance of minority and nonminority and women- and men-owned businesses (see Appendix A).

The primary shortcoming of the SBO, in terms of its value for producing statistics on business dynamics, is that it is carried out infrequently—once every five years. Because many new businesses emerge then fail quickly, this kind of information needs to be collected on a more frequent basis.

Recommendation 5: The Census Bureau's SBO should be conducted on an annual basis. The survey should include both a longitudinal component and a flexible, modular design that allows survey content to change over time. In addition, the Census Bureau should explore the possibility of creating a public-use (anonymized) SBO or a restricted access version of the data file.

The survey could be modified to include panel elements as well, perhaps in a manner similar to what is done in the Annual Survey of Manufacturers. This would facilitate measurement of the transitions that young and small firms make over their lifetimes. Finally, it would allow for flexibility in the type of questions asked over time by incorporating survey modules that differ with respect to content. For example, to minimize burden, one could

create modules on business finance, investment, and workforce training, among others, and cycle through them so that each is conducted periodically. The net result of such a program would be more detailed statistics about young small firms, provided on a more consistent basis, with overall better survey coverage than is currently available.

Implementing Recommendation 5 entails no conceptual hurdles; however, a more frequent survey would create new demands on resources and raise concerns about burden. It is possible that respondent burden associated with a more frequently conducted survey could be offset by rotating the samples and supplementing the survey data with additional administrative data. Finally, the value of the SBO would be greatly enhanced if researchers could obtain greater access to the microdata in secure settings or through creation of a public-use file.

5.2 MORE EFFECTIVE USE OF EXISTING INFORMATION

> Statistical agencies do not and should not conduct their activities in isolation. An effective statistical agency actively explores ways to work with other agencies to meet current information needs, for example, by seeking ways to integrate the designs of existing data systems to provide new or more useful data than a single system can provide. . . . Efforts to standardize concepts and definitions, such as those for industries and occupations, further contribute to effective coordination of statistical agency endeavors, as does the development of broad macro models such as the system of national accounts.
>
> —Practice 11: Coordination and Cooperation with Other Statistical Agencies, *Principles and Practices for a Federal Statistical Agency* (National Research Council, 2005a, p. 41)

In working toward an improved and more versatile data system, the question of how much and what kinds of data are needed to fulfill important purposes must be balanced against the cost and burden associated with the enterprise. Given finite, often tightening resources, a realistic strategy to improve business data must rely heavily on effective use of current data collection efforts.

5.2.1 Linking Survey and Administrative Data Sources

A comprehensive business data system must integrate information from an array of sources—private and public, business and household based, cross-sectional and longitudinal, survey and administrative, national and subnational—in a way that permits business dynamics to be measured in ways that are just now being conceptualized. Given legal, bureaucratic, and political realities, movement toward an ideal system can be expected to be

incremental, but the basic idea to creatively combine data sources should guide the process from the beginning.

A key aspect of the strategy to better coordinate data collection and production involves solving technical and legal hurdles so that administrative data that are routinely collected by (and from) businesses can be optimally exploited. Use of administrative data can (1) improve accuracy of information, particularly when survey questions require respondent recall, (2) improve breadth of information, and (3) reduce respondent burden by minimizing the amount of information that must be gathered using surveys (National Research Council, 2005a, p. 8). In conducting their business data programs, the statistical agencies could, if permitted, make more effective use of administrative data that are collected as a matter of course.

For studying topics related to life-cycle business dynamics—such as the link between the age of businesses and their economic contributions (e.g., to employment growth or innovation)—it is particularly important to develop a linking strategy that allows for the construction of comprehensive longitudinal data structures that capture events as they take place. Many of today's surveys and censuses have longitudinally incompatible questionnaires. When survey instruments are created and revised, more weight should be given to the potential longitudinal uses of the data. It would also be highly desirable to be able to link new collections to existing data sets to maximize their research and policy value. Linkage opportunities include tapping "nonbusiness" data, such as the CPS, the ACS, and the American Time Use Survey.

Looking forward, the statistical agencies should develop their administrative data and surveys with the intent to integrate them into a longitudinal household-business data infrastructure.

> **Recommendation 6: The Census Bureau should develop a fully integrated longitudinal household-business data infrastructure from administrative data to serve as a platform for tracking business formation, for integrating household and business survey data for measuring economic activity associated with the business formation process, and for developing samples for new surveys of business dynamics. The integration should include the master household address files, the job frame from linked employer-employee administrative records, and data for firms (including those with no paid employees, but with receipts) from the Census Bureau business register.**

The Federal Economics Statistics Advisory Committee recently advised BLS and the Census Bureau to further integrate household and employer data. One motivation is to investigate the discrepancies between the various employment statistics produced by the agencies. Given the potential differ-

ences in the treatment of young businesses and nonemployers in data originating from households versus businesses, measures of self-employment and, in turn, business formation and dynamics can be systematically affected. Again, implementing these recommendations will require little or no new data collection, just more intensive record processing.[2]

Moving toward the vision specified in this recommendation involves a long-term strategy. Elements of this strategy to innovatively integrate data and improve coverage of small and young firms are already in motion. The Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program combines federal and state administrative data on employers and employees with federal censuses and surveys. The LEHD has created opportunities to conduct research on topics for which empirical analysis of confidential longitudinal linked employer-household microdata are required. Similarly, the Integrated Longitudinal Business Database (ILBD) combines survey and administrative data on employer and nonemployer businesses. The ILBD provides a tool for studying business start-ups and early life-cycle dynamics by tracking business entities as they transition from nonemployer to employer status (Appendix A contains detailed descriptions of these data sources). Many other opportunities exist as well. For example, the SBO, discussed above, could be linked to the ILBD, which would allow for more thorough, though still imperfect, longitudinal tracking of businesses by owner(s)' race and gender.[3]

An efficient data collection infrastructure also requires that survey programs be well coordinated across statistical agencies. It is not sensible to expand the Census Bureau's surveys to include more information or to collect it at more frequent intervals when similar data are already collected in BLS surveys. Similarly, it is not efficient to add output and nonlabor input measurement (like capital investments) to BLS high-frequency surveys. However, periodic measurement of all these concepts on the same questionnaire (and from the same entities) is the only way to identify and correct errors in the estimation of dynamic relationships that occur when the microdata from multiple sources are aggregated for use in statistical products. Recognizing cost limitations, the panel recommends that the two agencies use topical modules in each other's surveys to address this defi-

[2]This is not to say that it will be easy. The quality and properties of administrative versus survey data, including the quality of the longitudinal links, should be a high-priority research topic in the development of this data infrastructure. There are also complications involving timing as it relates to linking survey and administrative data that are captured with different time lags.

[3]Something similar to this was done by Robb (2000) using 1992 business owner survey data and longitudinal establishment data.

ciency. The CPS is an example of a core statistical program that successfully uses a modular approach.

Recommendation 7: BLS and the Census Bureau should jointly develop intermittent topical modules for their business surveys. These topical modules should be designed to allow periodic measurement in the same survey and with the same business sample of variables usually collected in separate surveys and at different frequencies.

If survey data can be effectively combined and supplemented using other sources, then the amount of information that must be requested through more burdensome instruments and duplication of effort by respondents can be minimized. This would allow surveys to be used in a targeted way when detailed information on special topics is needed but not available from administrative data sources; examples include business start-up financing data from new firms and detailed demographic information on entrepreneurs. Small-scale surveys can also be used to corroborate the accuracy of administrative data.

Even targeted surveys do of course require resources but, if extensions can be built into existing programs, these outlays can be minimized. Our example of adding a supplement to the CPS for the self-employed, in order to generate more accurate statistics on the demographic characteristics of entrepreneurs, would require only a routine modification to the survey. Agencies should seek outside funding from foundations and the National Science Foundation to support these supplemental surveys.

Integrating multiple surveys and surveys with administrative and other kinds of data sources requires that clear and consistent definitions be used. This applies to the various measurement units (firms, establishments, sole proprietors, etc.); for tracking how these entities are born, grow, decline, and die; for defining boundaries and locations; and for identifying production and reporting units (employer and employee identifiers and industry and product codes). Consistency of terminology, concepts, and identifiers increases the likelihood that disparate data sources can be accurately linked.

Data sources should be modified to enhance the ability to link across government sources, as well as private-sector ones. In some cases, particularly for measuring activities of small and young businesses, a wealth of information is maintained in the private sector. The statistical agencies should try to take advantage of these resources.

Recommendation 8: The Census Bureau and BLS should explore and actively pursue opportunities to acquire microdata sets—on venture capital investment, business financing, and small business

lending—from commercial sources and from other government statistical agencies. Once acquired, these data sets should be integrated with existing business-level data sources at the Census Bureau and BLS to produce new public-release statistics on business activity by source and type of financing and to provide new tools for statistical analysis by qualified researchers.

Statistical agencies may be able to improve the accuracy and timeliness of existing data fields as well. Given that businesses must continually update their own employment, payroll, capital expenditure, and other records, it makes sense to develop conduits that allow internal reporting systems to feed into government data collections. By recognizing that companies maintain accounting systems associated with day-to-day operations on a high-frequency basis, they may, in the process, even be able to mitigate business respondent burden, particularly for large firms. There are obvious advantages when businesses can avoid having to repackage information for government surveys. New technologies (e.g., web-based reporting) will continue to enhance these kinds of opportunities to improve the timeliness and accuracy of collection efforts. For example:

Recommendation 9: The Office of Management and Budget should investigate the possibility of developing a common taxonomy, based on the extensible business reporting language, or XBRL (an industry-specific extensible markup language, XML), to allow common definitions to be used in surveys and administrative sources that can be automatically extracted from accounting and other business management software. In so doing, they should work with the statistical agencies, the Internal Revenue Service (IRS), accountancy organizations, and software providers. This will help meet the goals of paperwork reduction and may have applications for similar purposes beyond the statistical system.

XBRL, which has been described as the "digital language of business," provides the structure to reuse information in business reports. The Australian Bureau of Statistics has already stated that it sees XBRL as "likely to succeed as the industry accepted 'business reporting language.'"[4]

In terms of exploiting new and cost-effective data sources, options for integrating payroll data seem especially promising.

[4]See http://www.unece.org/stats/documents/2002/02/edr/20.e.pdf.

Recommendation 10: BLS and the Census Bureau should explore the possibility of continuous, real-time integration of payroll and employment data that are maintained by third parties into their systems; this could streamline data collection and, ultimately, possibly reduce respondent burden.

Recommendations 8, 9, and 10 are relatively low-cost ideas, given that the first steps are research-oriented. Setting up systems to actually use new kinds of data would require resources, but it could ultimately reduce costs and improve data timeliness. A firm like ADP—which processes payrolls for a client base that includes over half a million firms that account for 20 percent of private-sector employment—maintains a massive amount of data and even produces economic statistics.[5] Of course, exploiting this kind of data can never be more than one element of a comprehensive business data system.

Finally, surveys should be designed so that geographic location and sector of activity may be identified and aggregate statistics by industry, region, county, etc., constructed. The importance of tracking businesses and people at substate levels has been reinforced by the series of recent natural disasters that have disrupted and redirected many kinds of business and worker activity.

5.2.2 Coordinating the Business Lists

The business lists maintained by the statistical agencies serve a number of critical purposes. They are used to create sampling frames for a wide variety of surveys conducted by the Census Bureau and by other statistical agencies; for benchmarking survey data; for publishing employment and wage data; and for generating aggregates used by other agencies, most notably many of the inputs to the national income and product accounts. BLS and the Census Bureau have built extensive infrastructures designed to use administrative sources to create and maintain sampling frames that cover essentially the same universe: physical establishments with employees that engage in measurable economic activity.

Continued evolution of the U.S. business data system hinges, to a significant extent, on improving the Census and BLS business lists. The obvious way to make progress on this front is to exploit the best features of each, while reconciling inaccuracies and inconsistencies between them. Recent legislation—specifically the Confidential Information Protection and

[5]See, for example, the ADP National Employment Reports (http://www.adpemployment report.com/).

Statistical Efficiency Act of 2002 (CIPSEA), which allows sharing of confidential business data among BLS, the Bureau of Economic Analysis (BEA), and the Census Bureau for statistical purposes—provides a foundation facilitating the kind of data coordination needed to work toward this goal.

Recommendation 11: BLS and the Census Bureau should cooperate under the auspices of CIPSEA to initiate, and CIPSEA should be enhanced to allow, the creation and use by source agencies of a reconciled, consolidated, integrated business establishment list.

The potential advantages of an integrated business sampling frame are many: Census data would help BLS improve the consistency of industry codes and (if federal tax data can be shared) BLS could obtain information on the nonemployer universe, thereby improving sampling efficiency. BLS data could benefit the Census Bureau by providing employment data for single units, and industry codes and physical location information for all records. Matching BLS's business establishment list and the Census Bureau's business register would allow editing processes to be developed to identify records with large discrepancies. Gains from sharing the two existing business registers include more accurate measurements of births, deaths, and ownership changes; enhanced ability to track mergers and acquisitions; improved industry output and productivity measurement; and possibly reduced costs and burden. The potential benefits to downstream users from these upgrades (many of which are documented in National Research Council, 2006) are substantial.

Implementation of this recommendation would require harmonizing key elements of the current business lists. These elements include frame maintenance, activity/industry coding, birth and death record processing, ownership change identification, and handling of missing data. The creation of a reconciled, consolidated establishment register requires four broad steps involving information from the Census Bureau registers, BLS registers, and (indirectly) the state employment security offices, which maintain the input list used by BLS:

- matching and unduplicating the combined employer establishment lists of BLS and the Census Bureau (consolidated employer establishment list);
- integrating and unduplicating businesses from the Census Bureau's nonemployer registry with respect to establishments in the consolidated employer establishment list;
- validating the births, deaths, and entity demography using information from all sources; and

- reconciling the activity codes, physical locations, and volume variables when the consolidated list displays disagreements between the sources.

We envision that a reconciliation can be fully automated, as a cost-saving measure; however, human value added should be used to improve the process over time.

The new version of the establishment registry would include

- unique business identifiers,
- common activity codes (NAICS, presumably),
- common physical location identifiers (latitude, longitude, presumably),
- common volume variables (employment and payroll, presumably),
- common indicators of employer or nonemployer status, and
- common indicators of type of ownership and enterprise structure (e.g., multiunit).

Reconciliation of the two business lists in itself will not produce the kind of gains the panel envisions unless the most desirable characteristics of each can be brought to bear in the new product—reconciling the lists does not mean drawing from a single administrative and survey source. Even now, BLS and the Census Bureau should be sharing as much of the multiestablishment data (which does not involve IRS data) as they possibly can. A goal of the reconciliation project might be to identify the most productive data items and records to be shared under CIPSEA. To fully integrate all useful information, reconciliation should take place at or near the end of the production processes that yield the current BLS and Census Bureau business registers. An exception might be that new establishment lists be reconciled as soon as data become available, so that sample frames can be rapidly and continuously updated.

Finally, it is worth reemphasizing that business list comparison work is already well under way at the agencies. This progress not withstanding, the idea of a BLS-Census Bureau business list *reconciliation* is still very much in the discussion stages. The legalities and procedures necessary to get this started (most specifically, the restrictions resulting from tax return data that the Census Bureau receives from the IRS) are not trivial. A high proportion of Census Bureau business records contain data that originate or that are derived from tax records, the use of which is restricted by tax law and IRS regulations.

5.2.3 Expanding the Use of Data

Accurate and timely information about the economy is critical for effective policy making in both the private and public domains. The statistical agencies rightly view themselves primarily as data collectors, and their mission is to do so, maximizing quality subject to budget constraints. While the agencies have skilled in-house staffs of policy analysts and researchers, the vast share of expertise resides elsewhere, and public policy research must be done at universities or other nongovernmental institutions. It is this intensive use of statistical agency products that gives them their high public value.

Recommendation 12: The quality of research based on business data produced by the statistical agencies would improve with greater interaction between outside researchers and businesses and the statistical agencies. As recommended in previous Committee on National Statistics reports, statistical agencies, in particular the Census Bureau, should incorporate into their missions a broader interpretation of the criteria for access to data. Specifically, research that informs social and economic policy should be considered a valid reason for accessing confidential data.

The panel is encouraged by recent steps at the Census Bureau and IRS to emphasize the importance to public and private decision making of research that takes place at the agency's data centers, and to work out procedures to facilitate streamlined processes for reviewing proposed research projects.[6] It is also important that the statistical agencies facilitate front-end collaboration with the academic community (and, in some cases, the business community as well) with respect to survey design.

Ideally, the basic elements of the employer section of the business register should be made accessible to qualified researchers to serve as a sampling frame for their surveys. Practically, we realize that this will not happen overnight, and we recommend extending access first for key government policy research purposes.

[6]Our optimism is based, in part, on a January 4, 2007 memo from the Director of the Census Bureau which states: "United States Code Title 13 Chapter 5 directs the Census Bureau to carry out censuses and surveys of the U.S. population and economy. Ensuring that resulting data meet the highest standards of quality and utility requires significant supporting analytical research by Special Sworn Status researchers participating in the Census Research Data Center program. The importance of this research will only increase, as the data needs of public and private decision makers grow broader and more complex. . . . Accordingly, and to continue fulfilling its mandate at the highest level of technical excellence, it is the policy of the Census Bureau to undertake analytical research for authorized purposes, to the fullest possible extent."

Recommendation 13: It would be highly desirable if the business register(s) were available to federal agencies for the purpose of constructing sampling frames. For example, the Board of Governors of the Federal Reserve System should be able to access, in a secure setting that ensures current levels of confidentiality, the Census Bureau business register for the purpose of drawing samples for the surveys that they conduct or commission.

Changes in legislation protecting confidential information should permit sharing of data—including reconciled industry, location, entity identifier, volume, and employer status codes—for statistical purposes with the condition that the original source of the reconciled value and entity is not identified.

For many purposes—such as state and local planning—data must be collected and accessible in a way that allows for small-area analyses. Data with precise location identifiers are also needed to document the impact of federal government policies and actions. As a specific example, one could imagine research and policy interest in data on federal government contract awards to private businesses, linked to the Census Bureau business register using EINs or other common business identifiers. The Census Bureau could use data linked in this way to produce public-release statistics on the volume and type of contract awards by county, industry, business size, and business age. The impact on local economies of base closures, military reserve deployment, emergency relief, and extending eligibility for unemployment insurance are other examples of situations in which the capacity to fully analyze events has been compromised by data limitations.

The same attributes that make data useful for research and policy can also increase their value to businesses. For example, most business planning takes place at the subnational level, which generates a need for small-area statistics. However, additional burden can also be created since firms, especially multiunit firms, may have trouble disaggregating information into small geographic areas. Often, firms are able to report only at a more aggregate level than that which would truly be of interest. More generally, there are increasing conceptual difficulties associated with assigning a physical location to economic activity performed by businesses. It is becoming commonplace for economic activity to be conducted by "virtual businesses"—groups of people collaborating without a formal employment relationship. Globalization has also created more complex and far-flung supply chains.

The statistical agencies have done a good job of integrating small-area details in many of their data programs. Survey programs should continue to collect, and administrative record systems should maintain, data that enable (1) identification, for authorized purposes, of detailed geographic

and sectoral location of business activity, generally at the establishment level, and (2) flexible aggregation of statistics by product, industry, region, county, etc.

Geographic specificity raises confidentiality issues, since finer geocodes compromise anonymity. Typically, data at the finest level of geographic detail can be made available only in restricted access settings or, for public-use data sets, in an altered form. Agencies can add value to their products to the extent that they can utilize statistical "fuzzing" techniques to maintain confidentiality without completely losing geographic details.

Recommendation 14: Using synthetic data approaches or other statistical disclosure limitation techniques, BLS and the Census Bureau should work to develop anonymized, public-use versions of their recently developed longitudinal data sets on businesses. This should include the Longitudinal Database on Businesses from BLS and the Longitudinal Business Database and the ILDB at the Census Bureau.[7]

Work to begin implementing the recommendations in this subsection can begin without major cost outlays. These ideas (which are far from novel) involve no new data collection, only a fuller recognition of the public value of statistical data use and a shift in the policies regulating the scope of that use—the topic to which we now turn.

5.3 CHANGING THE DATA-SHARING ENVIRONMENT TO REALIZE SYSTEMIC EFFICIENCY

> Initiatives for sharing data among statistical agencies (including individual data and address lists when permitted by law and when sharing does not violate confidentiality promises) can be helpful for such purposes as achieving greater efficiency in drawing samples, evaluating completeness of population coverage, and reducing duplication among statistical programs, as well as reducing respondent burden.
>
> —Practice 11: Coordination and Cooperation with Other Statistical Agencies, *Principles and Practices for a Federal Statistical Agency* (National Research Council, 2005a, pp. 44-45)

We have emphasized a strategy for improving business data and statistics that involves more effective use of administrative and survey data that are already collected. Data sharing among the agencies is a key aspect of this idea. In order to produce the highest quality data sets and statistics at

[7]It should be noted that Reiter and Kinney (2006) are engaged in work to produce a partially synthetic version of the Longitudinal Business Database.

the lowest possible cost, the statistical agencies must be able to access the best information available, system-wide. Data sharing has the potential to reduce respondent burden as well which, along with assurances of confidentiality, may increase the likelihood that businesses respond to survey requests.

To their credit, the statistical agencies have recognized the potential gains from data sharing, and survey and administrative data on U.S. businesses are shared to some extent among BEA, BLS, and the Census Bureau for statistical purposes.[8] Recommendation 11, above, argues for extending CIPSEA to increase the flexibility with which information can be shared among statistical agencies for purposes of constructing a comprehensive business register and for designing special surveys. Sharing of business registers is essential in order to continue to improve the accuracy of measures of industry output, compensation, and productivity trends. It would permit the statistical agencies to keep abreast of the dynamic economy by producing statistical samples that are consistently and quickly adjusted to reflect entry and exit of new businesses. This is especially important for fast-growing and innovative industries, such as information technology. Such improvements would enhance our ability to perceive emerging trends in the economy and more accurately forecast economic activity.[9]

The panel endorses most aspects of the past efforts (reviewed in 4.4 Appendix) to expand data sharing. Effective coordination of statistical agency data programs is essential for improving the accuracy, coverage, and timeliness of business data, as well as the efficiency with which it is produced. As discussed above, a key part of the strategy to develop the most useful business data system possible (and a valuable and low-cost first step) would be to coordinate, and improve in other ways, the business lists residing at the Census Bureau and at BLS. Expanded interagency data sharing is a prerequisite for making progress on such a project.

Before work can progress much further to reconcile the business lists, and before data sharing between the three CIPSEA-designated agencies (BLS, BEA, and the Census Bureau) can be fully exploited, the IRS regulations and tax code legislation must be changed.

Recommendation 15: Measures should be taken immediately to facilitate the expansion of CIPSEA to increase the kinds of infor-

[8]The "statistical purposes" qualifier excludes using information for administrative, regulatory, law enforcement, judicial, or other purposes that may affect the rights, privileges, or benefits of a respondent.

[9]This argument was well articulated in comments made by Federal Reserve Board governor Randall S. Kroszner about "developing innovative statistics for a dynamic economy" (Kroszner, 2006).

mation that may be shared among statistical agencies for the purpose of reconciling the business lists and for the design of special surveys. This expansion of data sharing can be accomplished by (1) Congress acting to enact legislation that revises Internal Revenue Code Section 6103(j) to extend authorized access of IRS tax information to BEA and BLS, (2) the Treasury Department initiating an update of the IRS regulations, which clarify purpose and detail specific items that can be shared with authorized agencies, or (3) a combination of both actions.

While recognizing that the "ideal" data system for measuring business dynamics would ultimately integrate data from an array of sources—private and public, business- and household-based, cross-sectional and longitudinal, survey and administrative, national and subnational—it is worth noting that CIPSEA will, in reality, probably expand only incrementally. A first step should be a push to amend Section 6103 of the IRS Code and Treasury regulations to allow BLS and BEA access to part or all of the tax data to which the Census Bureau has access for the specific purposes of creating a unified business list. This might entail limiting data sharing, for federal tax information, to a small number of variables needed for business list coordination (e.g., name, address, legal form of organization, ownership structure, identity of parent firm if applicable, industrial classification, geographic coding information, employment, and payroll).

The goal of the agencies charged with creating and maintaining the source lists should always be to include accurate data on all business units, large and small, new and old. There are, according to the Census Bureau, roughly 18 million nonemployer firms[10]—about three quarters of all firms—and they constitute a reasonably large fraction (12 percent in 2000) of aggregate U.S. business revenues (http://www.census.gov/epcd/www/smallbus.html). In addition, for the subset of businesses in which this panel is particularly interested—the young and small ones, many of which operate in service and information sectors—there is much fluidity between those that have employees and those that do not. Indeed, a substantial portion of established firms in the United States started out as nonemployers (many of them sole proprietorships). Thus, sharing of business data would be quite limited if it did not permit an integration of the list information from the Census Bureau on nonemployer businesses.

[10]For nonemployers, a firm is the same as an establishment. Because the Census Bureau counts each distinct business income tax return filed as an establishment, it is possible for an individual to account for more than one nonemployer establishment.

Recommendation 16: In order to create a comprehensive business list and to generate data that would be useful for studying the dynamics of small and young firms, interagency sharing agreements should extend to data on nonemployers. Data on all sole proprietors and partnerships must also be included, whether they have employees or not.

We believe that a compelling data-driven case has been made, in this report and elsewhere, that reconciliation of the business lists would better serve downstream users—such as BEA in the production of national accounts and the Federal Reserve in carrying out research to inform monetary policy—to an extent that more than warrants the actions recommended here. In order to effect such changes, active support will be needed from within the administration (e.g., the Office of Management and Budget and the Council of Economic Advisers, where it appears to already exist) and from congressional staff. A key element to generating this support involves ensuring that further sharing of business data for statistical purposes does not unduly compromise confidentiality. The political and legal feasibility of such an extension has certainly been enhanced by recent events. The provisions of CIPSEA provide sufficient coverage to continue to ensure that the privacy and confidentiality of records will be maintained, even with expanded sharing of information. Indeed, given the uniform set of confidentiality requirements enacted through CIPSEA, the agencies are now in a better position than ever before to protect data collected for statistical purposes under a pledge of confidentiality.

Finally, one danger associated with improving data quality through cooperative arrangements is that it may increase the risks (or perceived risks) of allowing researcher access. Interagency data sharing efforts are clearly desirable, but precautions must be taken to ensure that the improved richness of the business data source resulting from linking, sharing, or better coordination of administrative and tax records cannot rightly be used as an argument for further restricting access.

5.4 RECOMMENDATION PRIORITIES AND COSTS

In this report, we have presented our views expressing (a) why the United States needs to improve its measures and understanding of business dynamics; (b) why this requires obtaining better data, especially longitudinal information, on new and small businesses than are currently produced; (c) what an ideal data collection system for monitoring business dynamics might look like; and (d) some of the steps that would need to be taken to facilitate an ongoing and feasible data collection effort that is sensitive to confidentiality, legal, and cost considerations.

In this chapter, we have presented our recommendations; however, we realize that this is a very broad outline, and creative work will have to continue within the statistical agencies, by outside researchers, and through collaborations between these groups. The statistical agencies should continue to tap into their advisory committees (such as the Federal Economic Statistics Advisory Committee) to provide further guidance on prioritizing the recommendations in this report as well as new ideas that will continue to emerge.

While the panel cannot provide detailed cost estimates of what the recommendations in this report would cost to implement, we have attempted to provide guidance on the topic, and we summarize some of our views on costs and priorities here.[11] Work could begin almost immediately to implement several of the panel's recommendations, and with modest resource commitments. Actions associated with two of the report's core themes could accurately be categorized as low-hanging fruit—they involve minimal long-term monetary outlay (there may be political costs) but are likely to yield high value: (1) The statistical agencies should *maintain* the business register in a more coordinated manner—Recommendations 11, 15, and 16—and (2) they should *utilize* it to produce new tabulations of economic activity by measures of business (establishment) age—Recommendations 2 and 3.

The specifics of point 1 involve expanding information sharing by government agencies and the use of such things as common identifiers to make it easier to link data across administrative and survey sources. The report provides guidance on how these objectives could be attained without unduly compromising confidentiality and without creating excessive new reporting burdens to businesses. Part of the strategy—for example, to develop consistent and comprehensive data on nonemployer businesses—involves more intensive and better coordinated use of existing administrative records. Implementing point 2 also seems to us quite feasible. Business age variables can be constructed by linking the business registry over time; the Census Bureau's Center for Economic Studies has already done this through its construction of the Longitudinal Business Database. Publishing information about how economic activity varies with establishment age is essentially a matter of producing new tabulations (some of this is already taking place with the SUSB program, which publishes total employment in establishments that are one year old or less). One potential cost embedded in this recommendation is that, when publishing tabulations, trade-offs must sometimes be confronted in order to maintain respondent confidentiality. So,

[11]Our organizational thinking for this section was shaped by the comments of two reviewers of the report.

while age is an important variable, the panel is not suggesting giving up equally and, for many applications, more important industry and geographic detail to acquire it.

The report also includes recommendations—specifically numbers 1, 4, and 5—that would be more expensive to implement and that would require a longer term concerted effort to carry out. Recommendation 1 states that young business should be sampled more in existing business surveys, which implies that either the number of large firms that are sampled must be reduced or that funding for the surveys must be increased marginally to cover costs associated with a larger survey. The rationale underlying this guidance is that young and small establishments should be given more weight than their receipts or total employment might suggest, because their characteristics change quickly and because they may contribute disproportionately to growth.

Recommendation 4 proposes incrementally adding questions to household surveys to identify and gather information about nascent entrepreneurs. The panel recognizes the inherent inefficiencies in using household surveys for this purpose. While the cost of creative integration of data sources and of adding a module to an existing survey is not exorbitant (the former could even lead to cost savings down the road), following firms from their genesis forward is not a trivial task. Since this is a potentially expensive undertaking, it is a lower priority, or at least a more long-term goal. Recommendation 5 proposes that the SBO be conducted on an annual basis rather than a quinquennial basis, which would require new funding. It should be noted that none of these recommendations suggest the creation of major new data collection efforts; most involved adjustments or reorganization of existing activities. Even the proposals for new data collection involve additions, enhancements, or modifications of existing administrative or survey efforts. In fact, almost all of the proposed additional procedures have already been developed and extensively field tested in one context or another.

Most of the report's remaining recommendations have to do with program processes and data access issues that are already within the realm of ongoing agency responsibilities. Other recommendations, such as those suggesting more rapid integration of new technologies or more effective use of existing data sources, are offered with the hope of encouraging long-term efficiency of business data collection.

In summary, a major justification for the panel's recommendations is to avoid the costs of "benefits foregone" from the absence of timely, precise data on the mechanisms by which the U.S. economy adapts and grows. Even the most liberal estimates of the financial costs associated with all the recommendations would be in the tens of millions of dollars, spread across several federal agencies. The return on such an investment would be an

improvement in the data required to understand the complexity of business dynamics in the United States and, possibly, a reduction in the likelihood of mistakes in national policy making. The cost of a minor misstep that reduces the growth of a 13 trillion dollar economy by even a small fraction is in the billions of dollars. By any calculus, the amount of resources required to provide a more timely, more accurate, and more complete description of U.S. business dynamics seems like a very good investment of public resources, yielding substantial benefits for future generations.

References

Abowd, J.M., P. Corbel, and F. Kramarz
1999 The entry and exit of workers and the growth of employment. *Review of Economics and Statistics* 81(2):170-187.

Abowd, J.M., J. Haltiwanger, and J. Lane
2004 Integrated longitudinal employee-employer data for the United States. *American Economic Review Papers and Proceedings* 94(May).

Abowd, J.M., F. Kramarz, and D.N. Margolis
1999 *Minimum Wages and Employment in France and the United States*. (NBER Working Paper #6996). Cambridge, MA: National Bureau of Economic Research.

Abowd, J.M., and S.D. Woodcock
2001 Disclosure limitation in longitudinal linked data. Pp. 215-278 in P. Doyle, J.I. Lane, J.J. Theeuwes, and L. V. Zayatz, eds., *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies* (Chapter 10). Amsterdam, Netherlands: North-Holland.

Armington, C.
2004 *Development of Business Data: Tracking Firm Counts, Growth, and Turnover by Size of Firms*. (Prepared for the SBA Office of Advocacy). Washington, DC: U.S. Small Business Administration.

Atkinson, T.
2005 *Atkinson Review: Final Report: Measurement of Government Output and Productivity for the National Accounts*. London, England: Palgrave Macmillan.

Baldwin, J.R., D. Beckstead, and A. Girard
2002 *The Importance of Entry to Canadian Manufacturing*. (STI Working Paper #2002/3). Paris, France: Organisation for Economic Co-operation and Development.

Bartelsman, E.J., and M. Doms
2000 Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature* 38(3):569-594.

Becker, R., J. Haltiwanger, R. Jarmin, S. Klimek, and D. Wilson
2006 Micro and macro data integration: The case of capital. In D.W. Jorgenson, J.S. Landefeld, and W.D. Nordhaus, eds., *A New Architecture for the National Accounts*. Chicago, IL: University of Chicago Press.

Benedetto, G., J. Haltiwanger, J. Lane, and K. McKinney
Forthcoming Using worker flows in the analysis of the firm. Submitted to *Journal of Business and Economic Statistics*.

Berney, R.
1999 Comments. Prepared for the Small Business Session, Understanding the Digital Economy Conference, May 25-26, Department of Commerce, Washington, DC. Available: http://www.technology.gov/digeconomy/powerpoint/berney/index. html.

Black, S.E., and L.M. Lynch
2005 Measuring organizational capital in the new economy. In C. Corrado, J. Haltiwanger, and D. Sichel, eds., *Measuring Capital in the New Economy*. Chicago, IL: University of Chicago Press.

Bureau of Labor Statistics
2006 *Labor Market Information Cooperative Agreement*. Washington, DC: Bureau of Labor Statistics.

Burgess, S., J. Lane, and D. Stevens
2001 Churning dynamics: An analysis of hires and separations at the employer level. *Labor Economics* (8):1-14.

Clayton, R., and J. Spletzer
2005 Business Employment Dynamics. Paper presented at the April NBER-CRIW Producer Dynamics: New Evidence from Micro Data. Available: http://www.nber.org/confer/2005/CRIWs05/clayton.pdf.

Corrado, C., C. Hulten, and D. Sichel
2006 *Intangible Capital and Economic Growth*. (NBER Working Paper #11948). Cambridge, MA: National Bureau of Economic Research.

Davenport, C., and the Administrative Conference of the United States
1976 Report on Administrative Procedures of the Internal Revenue Service, October 1975, to the Administrative Conference of the United States. Senate document, 94th Congress, 2nd Session; 94-266.

Davis, S.J., J.C. Haltiwanger, and S. Schuh
1996 *Job Creation and Destruction*. Cambridge, MA: MIT Press.

Davis, S.J., J. Haltiwanger, R.S. Jarmin, C.J. Krizan, J. Miranda, and A. Nucci
2006 *Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes*. (Center for Economic Studies Working Paper #06-04). Washington, DC: U.S. Census Bureau.

Doyle, P., J.L. Lane, J. Theeuwes, and L. Zayatz, eds.
2001 *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*. New York: Elsevier.

Dunne, T., M.J. Roberts, and L. Samuelson
1988 Patterns of firm entry and exit in U.S. manufacturing industries. *RAND Journal of Economics* 19(4):495-515.

Fairlie, R.W.
2004 Recent trends in ethnic and racial self-employment. *Small Business Economics* 23(3):203-218.
2006 *Kauffman Index of Entrepreneurial Activity: National Report 1996-2005*. Kansas City, MO: Ewing Marion Kauffman Foundation.

Financial Accounting Standards Board
1976 Financial Reporting for Segments of a Business Enterprise. In *Statement of Financial Accounting Standards No. 14*. Norwalk, CT: Financial Accounting Foundation.

Foster, L., J. Elvery, R. Becker, C. Krizan, S. Nguyen, and D. Talan
2005 A Comparison for the Business Registers Used by the Bureau of Labor Statistics and the U.S. Census Bureau. Paper presented at the American Statistical Society meeting, August 7, Minneapolis.

Foster, L., J. Haltiwanger, and C.J. Krizan
2001 Aggregate productivity growth: Lessons from microeconomic evidence. In E. Dean, M. Harper, and C. Hulten, eds., *New Directions in Productivity Analysis*. Chicago, IL: University of Chicago Press.
2002 *The Link Between Aggregate and Micro Productivity Growth: Evidence from Retail Trade*. (NBER Working Paper #W9120). Cambridge, MA: National Bureau of Economic Research. Also available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=325946.

Gartner, W.B., N.M. Carter, and P.D. Reynolds
2004 Business startup activities. Pp. 285-298 in W.B. Gartner, K.G. Shaver, N.M. Carter, and P.D. Reynolds, eds., *Handbook of Entrepreneurial Dynamics: The Process of Business Creation*. Thousand Oaks, CA: Sage Publications.

Greenia, N., J. Lane, and B. Jensen
2001 Business perceptions of confidentiality. Pp. 395-429 in P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.V. Zayatz, eds., *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam, Netherlands: North-Holland.

Haltiwanger, J., J. Lane, and J. Spletzer
1999 Productivity differences across employers: The roles of employer size, age, and human capital. *American Economic Review* 89(2):94-98.

Hampton, P.
2005 *Reducing Administrative Burdens: Effective Inspection and Enforcement*. London, England: HM Treasury. Also available: http://www.dti.gov.uk/files/file22988.pdf.

Holmes, T., and J. Schmitz
1995 On the turnover of business owners and business managers. *Journal of Political Economy* 103(5):1005-1038.

Jarmin, R., and J. Miranda
2002 *The Longitudinal Business Database*. (Economic Studies #02-17). Washington, DC: U.S. Census Bureau.

Jorgenson, D.J., S. Landefeld, and W.D. Nordhaus, eds.
2006 *A New Architecture for the U.S. National Accounts*. Chicago, IL: University of Chicago Press.

Kaplan, S.N., B.A. Sensoy, and P. Strömberg
2005 *What Are Firms? Evolution from Birth to Public Companies*. (NBER Working Paper Series #11581). Cambridge, MA: National Bureau of Economic Research.

Kroszner, R.S.
2006 Innovative Statistics for a Dynamic Economy. Remarks at the National Association for Business Economics Professional Development Seminar for Journalists, Washington, DC, May 24.

Loughran, D.S., J.A. Klerman, and B. Savych
2006 *The Effect of Reserve Activations and Active-Duty Deployments on Local Employment During the Global War on Terrorism*. Santa Monica, CA: RAND Corporation.

Madison, J.H.
1975 The credit reports of R.G. Dun & Co. as historical sources. *Historical Methods Newsletter* 8:128-131.

McGuckin, R.H.
1992 *The Importance of Establishment Data in Economic Research*. (Center for Economic Studies Discussion Paper). Washington, DC: U.S. Census Bureau.
1995 Establishment microdata for economic research and policy analysis: Looking beyond the aggregates. *Journal of Business and Economic Statistics* 13(1):121-126.

National Research Council
1993 *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Panel on Confidentiality and Data Access, Committee on National Statistics. G.T. Duncan, T.B. Jabine, and V.A. de Wolf, eds. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
1998 *Measuring the Government Sector of the U.S. Economic Accounts*. Committee on National Statistics. C.M. Slater and M.H. David, eds. Washington, DC: National Academy Press.
2000 *Improving Access to and Confidentiality of Research Data: Report of a Workshop*. Committee on National Statistics. C. Mackie and N. Bradburn, eds. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
2005a *Principles and Practices for a Federal Statistical Agency, Third Edition*. Committee on National Statistics. M.E. Martin, M.L. Straf, and C.F. Citro, eds. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
2005b *Expanding Access to Research Data: Reconciling Risks and Opportunities*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
2006 *Improving Business Statistics Through Interagency Data Sharing: Report of a Workshop*. C. Kuebler and C. Mackie, Rapporteurs. Steering Committee for the Workshop on the Benefits of Interagency Business Data Sharing. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Neumark, D., J. Zhang, and B. Wall
2005 *Employment Dynamics and Business Relocation: New Evidence from the National Establishment Time Series*. (NBER Working Paper #11647). Cambridge, MA: National Bureau of Economic Research.

Nichols, E., and D.K. Willimack
2001 Balancing Confidentiality and Burden Concerns in Censuses and Surveys of Large Businesses. Paper presented at the Federal Committee on Statistical Methodology Research Conference, November, Arlington, VA.

O'Donnell, G.
2004 *Financing Britain's Future: Review of Revenue Departments*. London, England: HM Treasury. Also available: http://www.hm-easury.gov.uk/media/149/63/odonnell_fore_ ch1_245[1].pdf.

Office for National Statistics
2005 Compliance Plan 2005-06 to 2007-08. Available: http://www.statistics.gov.uk/downloads/theme_other/Complianceplan200506to200708.pdf.

Ohlsson, E.
1995 Coordination of samples using permanent random numbers. Pp 153-169 in B. Cox et al., eds., *Business Survey Methods*. New York: Wiley.

Okolie, C.
2004 Why size class methodology matters in analyses of net and gross job flows. *Monthly Labor Review* July:3-12. Also available: http://www.bls.gov/opub/mlr/2004/07/art1full.pdf.

Pierce, K.
2005 Sole proprietorship returns, 2003. *Statistics of Income Bulletin June 22*. Available: http://www.irs.gov/pub/irs-soi/03solp.pdf.

Pilat, D.
2002 Work on Enterprise Demography at the OECD: An Update. Paper presented at 16th International Roundtable on Business Survey Frames, Lisbon, October 21-25. Available: http://www.ine.pt/novidades/semin/round_table/papers/paper/OECD_ Paper.doc.

Raghunathan, T., J. Reiter, and D. Rubin
2003 Multiple imputation for statistical disclosure limitation. *Journal of Official Statistics* 19(1):1-16.

Reiter, J., and S. Kinney
2006 Making a Public Use, Synthetic Version of the Longitudinal Business Database. Paper presented at the International Comparative Analysis of Enterprise (Micro) Data Conference, Chicago.

Reynolds, P.D.
2007 *New Firm Creation in the U.S.: A PSED I Overview*. Cheltenham, England: Edward Elgar.

Reynolds, P.D., N.S. Bosma, E. Autio, S. Hunt, N. De Bono, I. Servais, P. Lopez-Garcia, and N. Chin
2005 Global entrepreneurship monitor: Data collection design and implementation: 1998-2003. *Small Business Economics* (24):205-231.

Robb, A.
2000 *The Role of Race, Gender, and Discrimination in Business Survival*. Ph.D. dissertation. Ann Arbor, MI: University of Michigan Press.

Rubin, D.B.
1993 Satisfying confidentiality constraints through the use of synthetic multiply-imputed microdata. *Journal of Official Statistics* 9:461-468.

Salamon, L., and S.W. Sokolowski
2005 Nonprofit organizations: New insights from QCEW data. *Monthly Labor Review* (September):19-26.

Spletzer, J.R., R.J. Faberman, A. Sadeghi, D.M. Talan, and R.L. Clayton
2004 Business employment dynamics: New data on gross job gains and losses. *Monthly Labor Review* (April):29-42.

Syverson, C.
2004 Market structure and productivity: A concrete example. *Journal of Political Economy* 112(6):1181-1222.

Tehan, R.
2003 E-commerce Statistics: Explanation and Sources. Report for Congress. Congressional Research Service, Library of Congress, June 4.

U.S. Congress
2002 Confidential Information Protection and Statistical Efficiency Act of 2002. (P.L. 107-347).

Vale, S.
2006 *The International Comparability of Business Start-up Rates, Final Report*. Paris, France: Organisation for Economic Co-operation and Development, November. Available: http://www.olis.oecd.org.

Van Wey, L.K., R.R. Rindfuss, M.P. Gutmann, B. Entwisle, and D. Balk
2005 Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Science*s 102(43):15337-15342.

Willimack, D.K,. and E. Nichols
2001 Building an alternative response process model for business surveys. In *Proceedings of the Annual Meeting of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association.

Willimack, D.K., E. Nichols, and S. Sudman
2002 Understanding unit and item nonresponse in business surveys. Pp. 213-228 in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, eds., *Survey Nonresponse*. New York: Wiley.

Appendix A

Overview of Current Data Collections

This appendix provides an overview of current data sources used to construct business and employment statistics and to inform research and policy related to business formation, dynamics, and performance. Our focus is on data produced by the U.S. federal statistical system, but we also cite other examples.

The material in this appendix is organized into subsections loosely defined in terms of data source characteristics and purpose:

- data to count firms and catalogue essential characteristics—the business lists;
- longitudinal data for tracking businesses over time;
- data sources designed to improve coverage of small businesses;
- aggregate employment statistics;
- data on the self-employed, entrepreneurs, and business gestation;
- coverage of special sectors, such as agriculture, nonprofit organizations, and e-commerce; and
- financial data.

Beyond describing the basic design elements, we indicate the extent to which data are available to users outside the agencies (or other organizations) that collect them. Statistical agencies generally provide documenta-

tion for their accessible data sources, and we try to avoid reproducing detailed descriptions that can be accessed elsewhere.[1]

Table A-1, located at the end of this appendix, allows quick cross-comparisons of various data sets. In this appendix, we omit several important kinds of business data that are more closely linked to production of aggregate statistics and are less central to the panel's charge. For example, we do not directly discuss price data—notably the producer price index (PPI), which measures changes over time in the selling prices received by producers of goods and services—or the array of industry and input/output data (much of it deflated by PPI) crucial to productivity measurement and to the construction of the national accounts and statistics on gross domestic product (GDP).

A.1 COUNTING FIRMS AND CATALOGING ESSENTIAL CHARACTERISTICS—THE BUSINESS LISTS

The two primary business lists administered by federal statistical agencies in the United States are the Census Bureau's business register (BR), and the Bureau of Labor Statistics' Quarterly Unemployment Insurance (UI) address file, more commonly referred to as the Quarterly Census of Employment and Wages (QCEW). Administrative data from the Internal Revenue Service (IRS), which maintains the Business Master File, and the Social Security Administration (SSA) underpin the BR, while the QCEW relies on data from the state UI programs. The most noteworthy business list maintained outside government is the Dun & Bradstreet Dun's Market Identifiers (DMI).

A.1.1 The Census Business Register

In 1968, the Office of Management and Budget directed the Census Bureau to develop and maintain a comprehensive business list. Known until recently as the Census Bureau's Standard Statistical Establishment List (SSEL), the BR covers the universe of businesses—over 7 million employer businesses and some 16.5 million nonemployer businesses. The BR serves as the master enumeration list for sampling frames drawn for the Census Bureau's firm and establishment surveys, most notably the quinquennial

[1]The Kauffman Foundation web page (http://research.kauffman.org) has a well-organized list, with links of government and private sources of data on U.S. and international businesses; the list focuses on entrepreneurship, small business, and self-employment information. RAND, with funding from the Kauffman Foundation, has also assembled an overview of data resources on small businesses—see http://www.rand.org/pubs/working_papers/WR293/index.html.

economic census. The economic census, conducted during years ending in 2 and 7, covers over 5 million companies; nonemployer and small businesses are covered by sample only, not a full census (http://www.census.gov/econ/overview/mu0000.html). Domestic, nonfarm business data are collected at the metropolitan statistical area (MSA) geographic level. Because it occurs only every five years and a firm can materialize and close (or vice versa) over shorter periods, the economic census does not comprehensively capture business birth and death information.

The BR also serves the important function of providing central storage for an array of administrative data—most notably, payroll tax records, corporate and individual tax returns, and Employer Identification Number (EIN) application information. Maintenance of the BR is heavily dependent on these administrative data. Data on nonemployer firms are drawn exclusively from administrative sources, mainly business income tax records.[2]

Within the BR, data are organized at the establishment level—that is, a single location where goods are produced or services provided. Reflecting the composition of the economy, most are single-unit businesses, but there are establishments that are part of businesses operating in multiple locations as well. Because taxes—and in turn tax information—are collected from firms, Census researchers must break up IRS administrative data to the establishment level for multiunit enterprises.[3] In interim business census years, this is done using information from the Company Organization Survey (COS)—an annual survey of all large employers (250 employees or more) and a sample of smaller mid-size companies, reaching approximately 50,000 of the largest multiunit enterprises. The accuracy of the single/multiunit identification is reported to improve around economic census years, and then to decline thereafter (Jarmin and Miranda, 2002). The COS is used more generally in an attempt to maintain up-to-date company affiliation, location, closings, spin-offs, and operating information for multiestablishment companies. This allows for fuller coverage of such companies, which account for the vast majority of the nation's business activity. Title 13 of U.S. Code authorizes this and the other economic census-related surveys and stipulates mandatory responses.

A key element of the BR program is the identification and tracking of individual establishments owned by multiestablishment firms. The BR has

[2]See Jarmin and Miranda (2002) for a thorough description of the Census Bureau's business register, including specifics about industry-level coverage.

[3]As noted throughout this report, unit of observation is a key issue. Administrative data, such as those originating with the IRS, are collected and stored by taxpayer ID. The Census Bureau and BLS create enterprise and establishment units through supplemental surveys and various processing techniques. Both the QCEW and the BR are establishment-based but built from data organized at the taxpayer ID (EIN) level.

excellent coverage of multiestablishment firms every five years, but in the years between the economic census, the Census Bureau relies on the COS to update the multiunit segment of the business registers along with information it learns about multiestablishment firms from its other surveys (e.g., Annual Survey of Manufactures[4]). The limited scope of the sampled firms and the rotation of these sampled firms over time affect the timeliness and coverage of smaller multiestablishment companies in the business register. Several studies have noted that both births and deaths of smaller multiunit establishments are more concentrated in the year prior to the economic census when the register is being prepared for the upcoming economic census.[5] One sees this pattern of the clumping of changes in other dimensions of the data as well. McGuckin and Peck (1992) show that industry coding changes for establishments were especially concentrated in the economic census years.

A number of improvements were made to the newest version of the BR, which became fully operational in January 2002: additional data elements were added; seven years of data are now maintained, instead of three, allowing tracking of businesses from one quinquennial census to the next; processing of nonemployer statistics, which previously were not maintained, has been expedited; and industry detail has been brought into concordance with the North American Industry Classification System (NAICS). In addition, in 2005, the IRS began providing quarterly employment data from tax form 941 instead of only for the first quarter. Form 941 includes the EIN, employer-reported wages and other compensation, employment for the pay period, income and social security tax withholdings, and related information. When a new business payroll record is received from the IRS, the Census Bureau adds a business employer record to the BR. Nonemployers cannot be identified as quickly, since personal income tax returns are filed annually rather than quarterly. Form 941 now also includes an identifier for businesses filing final tax returns—useful for capturing business deaths. In July 2004, Census began receiving SS-4 form data directly from the IRS (rather than by way of SSA, as before) on a weekly basis, which allows industry codes to be assigned to new businesses more quickly.[6]

---

[4]For noneconomic census years, the Annual Survey of Manufactures provides sample estimates of employment, plant hours, payroll, number of establishments, cost of materials, value of shipments, inventories, and detailed capital expenditures statistics for commercial manufacturing establishments with paid employees (http://factfinder.census.gov).

[5]Jarmin and Miranda (2002) discuss the importance of retiming both small multiunit births and deaths in these data in order to improve the accuracy of the annual birth and death statistics.

[6]Salyers (2004) provides a "progress report" for the BR, including a full description of the expanded use of administrative records, as a more general listing of recent changes and improvements (http://www.stats.gov.cn/english/18roundtable/papers/t20041230_402219768.htm).

Data from the BR, as well as the more than 100 surveys that rely on its sampling frame, are used in the production of a wide range of publicly available aggregate statistics (many available on the Census Bureau's American Fact Finder web page at http://factfinder.census.gov). A widely used product of the BR is the Census Bureau's *County Business Patterns*. First-quarter employment and payroll numbers, cross-tabbed by county and kind of business, are published, cooperatively with the SSA, in the *County Business Patterns* and in the ZIP Code Business Patterns statistical series.

In addition, the Census Bureau's Non-Employer Statistics (NES) "provides U.S. and sub-national data by industry for businesses without paid employees." Originating primarily from administrative records, the NES "summarizes the number of establishments and receipts of sole proprietorships, partnerships, and corporations without paid employees." The Census Bureau began publishing NES data annually in 1997, and annual releases beginning with the year 2002 can be found on its American FactFinder web page.

These publications provide geographic aggregates of the BR microdata. BR data are also essential to economic research conducted at the Center for Economic Studies (for a description of these uses, see http://www.ces.census.gov/index.php/ces/1.00/researchprogram). Although a number of BR-based aggregate statistics enjoy high visibility, the BR is also structured with confidentiality very much at the fore. The BR itself is not a publicly available document, although parts of the register can be used by researchers under highly restrictive arrangements at the Census Bureau's research data centers (RDCs). Beyond this, data from administrative records are maintained in separate tables, and IRS Title 26 data are segregated from Census collected data. Microdata on race and gender, required for the Survey of Business Owners (SBO), is likewise stored in a separate table for use by SBO analysts only.

A.1.2 The BLS Business List

The other primary business list maintained in the federal statistical system is BLS's QCEW—formerly the Business Establishment List, initiated in 1988. The QCEW converts data submitted by the universe of employer businesses covered by state UI systems (ES-202), as well as federal agencies subject to the Unemployment Compensation for Federal Employees program, to an establishment basis. The master file includes a number of key fields: establishment name, address, telephone number, monthly employment and quarterly wages, federal EIN—all available by NAICS code, county, and ownership sector for the entire United States.[7] UI wage records

[7]Full details are documented at the BLS QCEW home page (http://www.bls.gov/cew/).

for individuals working in UI-covered employment are used at times by BLS and the states to validate individual cases of large wage fluctuations and include name, Social Security number, employer name and address, employer ID, and total earnings paid.

The QCEW serves as the sampling frame for most BLS surveys, and it is used to benchmark the Current Employment Statistics (CES) establishment survey. The establishment count also sets the population base in establishment birth and death estimators. The QCEW program provides a comprehensive source of employment and wage statistics, as well as a virtual census (98 percent) of employees on nonfarm payrolls (Spletzer et al., 2004). A crucial limitation of the QCEW—particularly in the context of understanding new and young business dynamics—is that it excludes nonemployer businesses and data on owner characteristics. The QCEW, which currently is geocoded to the rooftop level for 90 percent of private-sector employment, has plans for developing data at the census tract level. QCEW provides industry, employment, county, and physical location addresses on over 3 million firms, mostly new and small businesses, to the Census Bureau. However, the QCEW and BR have different structures which makes cross-survey comparisons difficult. In addition, requirements under the UI program's Multiple Worksite Report (MWR) vary from state to state and have size thresholds that may exclude certain businesses.[8] Finally, the ability to make longitudinal and cross-state linkages is complicated because no firm ID fields other than the tax ID number exist in the database (this is discussed again in the next section).

In the MWR, "multi-location employers with a total of 10 or more employees in their secondary locations are required or requested" to break out their employment and payroll by individual establishment. The MWR is mandatory in 21 states and provides good coverage for all but the smallest multiestablishment employers on a timely basis.[9] The timing of small multiestablishment births may not be accurate because reporting will depend on the secondary establishment passing a threshold size. Thus, when a single-location firm expands to a multilocation firm, one will not observe the "new" establishment until the establishment has at least 10 employees. In addition, if the expansion occurs across state lines, it may not be captured as a multiestablishment birth but as a new firm in the other state if it did not already have a presence in that state and if the firm has different EINs across states. There are also issues about firms that have multiple UI

[8]For example, businesses are not required to report a location in another state if there is only one, other sites within the state if total employment from these sites is less than 10; or any site that is under a different UI account number (http://bls.gov/cew/).

[9]BLS, http://www.bls.gov/cew/cewmrr00.htm.

and EIN accounts within a state that may affect multiestablishment measurement. QCEW has tried to identify across-state expansions in two ways. First, the state staff may notice a significant change in employment and wages reported by a firm. Upon follow-up, the staff may determine that a firm should file the MWR. If the change in employment and wages is small enough that the state staff does not observe the differences, the need for the MWR filing is captured after the employer completes the Annual Refiling Survey (ARS) and reports a new location.[10] About 2 million businesses are contacted annually to update such information as business name, address, and industry codes through the ARS.

As with the BR, numerous data products and statistics are derived from the QCEW, most prominently the quarterly wage and employment statistics, aggregated at various industry and geographic levels. Microdata underlying the QCEW are not publicly accessible; however, BLS does offer limited opportunities for researchers to access confidential data for the purpose of conducting statistical analyses. Data access is restricted to onsite use at the BLS national office in Washington (a list of the restricted access data sets available to researchers can be found on the BLS web site, http://www.stats.bls.gov/bls/blsresda.htm).

A.1.3 Dun's Market Identifiers

Business data are also collected by private-sector firms. These efforts are typically geared more toward marketing or informing business decisions and less toward research and public policy. The most prominent private-sector collection (and one that has been used for both purposes) is the Dun & Bradstreet (D&B) DMI. Because the BLS and Census business lists are not typically available as sampling frames outside those agencies, D&B data—and its Data Universal Numbering System (DUNS)—have been widely used in a variety of applications elsewhere in government. For example, it serves as the sampling frame for the Federal Reserve's Survey of Small Business Finances. The DUNS numbers are also used by the federal government to identify entities receiving federal contracts. Data have been broadly used by private-sector firms to estimate numbers of businesses, establishments, and employees, as well as sales and to perform cost-benefit analyses and risk assessment exercises. D&B data products can be purchased and used subject to the company's terms and conditions, which differ for end users (individuals, businesses, and information professionals).[11]

[10]Based on correspondence from Jim Spletzer, BLS.

[11]A full description of these terms and conditions can be found at http://library.dialog.com/bluesheets/htmlaa/bl0518.html.

DMI includes basic data, updated monthly, on over 2.9 million private and public companies and 17 million U.S. business establishment locations (about 18.4 million records as of January 2006) operating in private, public, and government spheres (there are also European and other international versions). The data set is broadly representative of all businesses but limited to private and public companies with five or more employees or sales of $1 million, and consequently, it does not include many of the newest start-up firms or self-employed individuals.[12] In contrast, the IRS reports that for 2003 about "19.7 million individual income tax returns reported nonfarm sole proprietorships" (Pierce, 2005), of which about 3 million filed a Schedule C-EZ, on which annual receipts totaled less than $25,000 (www.irs.gov). The file contains up to three years of basic data (the length of coverage varies by company), such as type of business, legal and trade names, physical and mailing addresses, geographical descriptions, product and industry descriptors, sales and number of employees (and the number at each corporate location), growth rates, annual sales, net worth and profit, names and titles of key executives, corporate linkages, DUNS numbers, and other marketing information.

D&B data are collected from various sources, such as in-person and telephone interviews, government publications, and business trade programs and mailings, a fact that limits the quality of information in some important ways. For example, there is no standard guideline for detecting new businesses and incorporating them into the file—information is brought in ad hoc from applications for credit, classified advertising, and other private sources. Similarly, there is no clear process for purging records. Unlike several of the government data sources, DMI does not have a mechanism for determining establishment versus firm records. Furthermore, the data are not longitudinal; in fact, DMI is not cross-sectional for a specific point in time, since there is no regular schedule for updates—the process is ongoing (Haviland and Savych, 2005).

A.2 TRACKING BUSINESSES OVER TIME: BUSINESS LIST-BASED SOURCES OF LONGITUDINAL MICRODATA

Sources of longitudinal business microdata have historically been scarce, particularly for smaller and newer businesses. However, new data programs are emerging that greatly enhance available information relevant to the topics covered in this report. Among the most promising data sets now or soon to be coming online are the Integrated Longitudinal Business Data-

[12]As such, D&B data have limited coverage of nonemployers.

base (ILBD) and Longitudinal Employer-Household Dynamics programs at the Census Bureau and BLS's Business Employment Dynamics.

A.2.1 ILBD and Precursors

The ILBD has evolved as a natural extension of the Longitudinal Business Database (LBD), which the Census Bureau's Center for Economic Studies began constructing in 1999. The LBD covers employer establishments, currently for the period 1975-2003. These programs, which can be traced to the early 1990s (under various names), have expanded research capabilities to new frontiers that would not have been possible with aggregate and cross-sectional data alone.

The LBD was constructed using EINs to link year-to-year snapshots of all employer establishments, along with name and location information contained in the Census Bureau's SSEL. Work is ongoing to add such fields as payroll employment, location, industrial activity, and firm affiliation. The LBD is useful for researching elements of business dynamics, such as firm entry and exit and job flows. Establishment identifiers also facilitate linking the LBD to other data sets. The value of the data set is enhanced by its algorithm for flagging establishment records as births, deaths, or continuers. Generally speaking, a birth is identified when a record appears for the current year that does not match any record from the previous year; a death is detected when a record for a previous year does not match any record for the current year; and continuing establishments show a match from one year to the next (see Jarmin and Miranda, 2002, for a detailed explanation of this algorithm). The practice of using EINs in conjunction with name and address information is intended to increase the accuracy with which establishment births and deaths can be identified; missing source data for some years make this a challenge.

The LBD itself is an extension of another CES predecessor, the Longitudinal Research Database (LRD), which contains longitudinally linked plant-level data from censuses and annual surveys of manufactures. With the relatively rapid growth and subsequent interest in nonmanufacturing sectors, the narrow focus of the LRD has become an increasing concern. Furthermore, LRD coverage of firms with fewer than 250 employees is limited, and the plant-level data are not linked to enterprises, so the overall size and industry of enterprises owning large plants are not always known. Despite these limitations, the LRD has been intensely analyzed and has spawned a robust literature (see Bartelsman and Doms, 2000, for a review of these efforts). The LBD allowed academic research on employment dynamics issues at the establishment level (forged by Dunne, Roberts, and Samuelson, 1989; Davis and Haltiwanger, 1990, 1992; and Davis,

Haltiwanger, and Schuh, 1996) to begin expanding beyond the manufacturing sectors.

The ILBD marks another discrete advance for business research aimed at understanding the processes of small and young firms over time, as its coverage is much broader than its predecessors. Extending work by researchers such as Boden and Nucci (2004),[13] the ILBD integrates federal government administrative records and survey sources for nearly all private, nonagricultural employer *and nonemployer* businesses in the United States, currently covering the years 1992 and 1994-2000 (see Jarmin and Miranda, 2003, Miranda et al., 2005, for a detailed description of the ILDB). One clear advantage of the ILBD over earlier data sets in the lineage is that it allows analysts to track a business's characteristics as it transitions from nonemployer to employer status (or vice versa), a key but difficult-to-study aspect of business evolution. ILBD data have shown, for example, that over three-year horizons, about 5 percent of nonemployer businesses become employer businesses or are acquired by, or absorbed into, employer businesses. This translates to approximately 750,000 businesses—a large number in absolute terms—and is an important component of job creation.

Employer businesses and some nonemployers are linked from period to period by EIN; most nonemployers are linked using business owner ID (Social Security number) fields. This technique is not seamless. For example, over time, ID numbers can change for legal or other reasons.[14] In addition, problems of inconsistent data formats, the volatility of young and small firms, and the sheer number of records (over 15 million nonemployers and over 5 million employer businesses) all pose challenges for the Census Bureau staff carrying out the project.

The ILBD has continually been under development and is not currently available to users outside the Census Bureau. Initial versions of some statistics are scheduled to be made available in the near future. Access to microdata will become available at RDCs, after further documentation of data quality assurance is completed, by perhaps as early as 2007. Access to ILBD data is governed by U.S. Code Title 13 (i.e., for statistical purposes only and with "predominant purpose" consistent with Census).[15]

---

[13]Richard Boden and Alfred Nucci linked nonemployer entities to the business register both cross-sectionally and longitudinally for the years 1992 through 1999. Their paper enumerates the myriad of issues that arise when attempting to track nonemployer businesses over time, including those involving sole proprietorships (not the least of which is a change in legal form of organization).

[14]The technical challenges inherent in ILBD linking procedures are documented in Davis et al. (2006).

[15]Documentation of the Census Bureau's RDC guidelines define "predominant purpose" and describes Title 13 requirements generally (http://www.ces.census.gov/index.php/ces/1.00/researchguidelines).

A.2.2 BLS's Business Employment Dynamics (BED) Program

The BLS's BED program produces a quarterly series of gross job gains and gross job losses statistics based on the universe of establishments covered in the QCEW (those subject to state unemployment insurance laws). Sectoral designations now conform to the NAICS classification system. Again, the major exclusions are the self-employed, along with certain nonprofit organizations. Data from the program were first published in September 2003 and are now complete for the period 1992 to the first quarter of 2006. Quarterly data will be released every three months, making them more timely than the alternative employment data sources previously available.[16]

The BED data allow disaggregation of employment changes into the underlying components—the number (and percentage) of gross jobs gained by opening and expanding establishments and the number (and percentage) of gross jobs lost by closing and contracting establishments.[17] These data, constructed using a multistep procedure to link QCEW microdata across periods, provide a picture of the dynamics underlying aggregate employment growth statistics.[18] Research based on the quarterly time series contributes to knowledge of the processes underlying the business cycle; for example, Clayton, Sadeghi, and Talan (2005) identify seasonally adjusted job changes resulting from establishment openings and closings, as opposed to expansions and contractions. In general, BED data have revealed that firm and establishment growth rates vary by size and how these results differ from those produced by analyses limited to annual data.

The primary obstacle to further development of the BED is that EINs are imperfect for creating record linkages (see Okolie, 2004); however both the QCEW and the BED incorporate a complex multi-stage process to link records across quarters.[19] As with QCEW microdata, researchers must submit proposals to access BED data; if the proposal is accepted, the data must be used at the BLS research center in Washington.

[16]These and other details can be found at http://www.bls.gov/bdm/bdmover.htm.

[17]Getz et al. (2005) provide a detailed description of the methodologies used to capture business births and deaths in the various Census Bureau and BLS data sources.

[18]Pivetz et al. (2001) describe the technique used to longitudinally link the data.

[19]Clayton, Sadeghi, and Talan (2005) provides some detail on the linkage procedures for the QCEW.

A.2.3 The Longitudinal Employer-Household Dynamics (LEHD) Program

The LEHD is a relatively new microdata source being developed by the Census Bureau, which describes the LEHD as a "set of infrastructure files using administrative data provided by state agencies, enhanced with information culled from demographic and economic (business) surveys and censuses. . . . The LEHD Infrastructure Files provide a detailed and comprehensive picture of workers, employers, and their interaction in the U.S. economy" (http://lehd.dsd.census.gov).

The program is breaking new ground by, as its name suggests, integrating data on households and individuals with data on employers. The idea behind the LEHD originated with a 1999 National Science Foundation initiative to investigate the potential to combine large administrative data sets with data collected through censuses and surveys. The objective was to "reduce respondent burden, increase data quality, and enhance the information available to the federal, state and local agencies which rely on Census Bureau data for decision making." The principal investigators on the project proposed linking information to permit data sets to be longitudinal in both the household/individual and firm/establishment dimensions (http://lehd.dsd.census.gov/led/about-us/FAQ.html#slehd).

The LEHD relies on BLS's QCEW in order to crosswalk between unemployment insurance accounts and the federal EINs and for its framework of detail on industry codes, employment levels, and physical location addresses. The LEHD is composed of interrelated infrastructure data sets: (1) the Employer Characteristic File, with information about the employer, including employment, payroll, industry, size, and location (http://lehd.dsd.census.gov/led/library/tech_user_guides.html); (2) the Employment History File, with information about the employment history of the employee, including employer identity, payroll, and employment; (3) the Personal Characteristics File, with time invariant information about the employee, including gender, race, foreign-born status, and date of birth; (4) the Employer Human Capital File, with statistics about the skill mix of businesses; and (5) the Employer Quarterly Workforce Indicators, with information at the employer level about accessions, separations, job creation, and destruction. These data sets allow for the integration of Census economic data with employee characteristics files (http://www.icpsr.umich.edu/access/census-unpub.html).[20] Core employee data originate from the QCEW. These records are supplemented with additional information on

[20]Full description of the LEHD structure can be found in Abowd, Haltiwanger, and Lane (2004).

individual and firm characteristics. The resulting database contains about 80 million individual and 5 million business records from participating states. Using these data, it is possible to follow each employer-employee match on a quarterly basis.

The LEHD has created opportunities to conduct research on topics for which empirical analysis of confidential longitudinal linked employer-household microdata are required, such as research on low-wage workers and human capital and productivity. The LEHD is already facilitating the creation of new statistics describing the dynamic nature of local economies. For example, it has spawned the Local Employment Dynamics program, a voluntary partnership between state labor market information agencies and the Census Bureau to develop new information about local labor market conditions. By receiving and processing quarterly data from each of about 29 state partners, quarterly workforce indicators are being produced by industry, age, and sex for local areas. The Quarterly Workforce Indicators program generates timely statistics on job churning, such as rates of accession, separation, job creation, and job destruction by detailed industry and location. Among the interesting results: Accession and separation rates have been found to exceed 20 percent per quarter, while rates of job creation and destruction are typically 7-10 percent. These statistics, which are comparable to the job creation and destruction rates from the BED, translate into over 13 million jobs destroyed each year (http://www.bos.frb.org/economic/ppb/2004/ppb0401.htm).

The ILBD integration of longitudinal data (survey and administrative) for all employer and nonemployer businesses has created a tool for studying business start-ups and early life-cycle dynamics (Davis et al., 2006). By incorporating geographical information system applications, analysts have been able to describe how workers travel to and from work for transportation planning purposes. One finding from this work, which has focused on workers leaving businesses (these separations and accessions are highly visible but account for only about 1 percent of the total), is that clusters of workers affected by outsourcing often move to temporary help and personnel supply jobs, helping to explain the growth of that industry (see Benedetto et al., 2004). The LEHD illustrates the tip of the iceberg in terms of the information volume and detail that can be made available through data integration and the efficiency of the approach relative to developing new surveys.

Data in the LEHD are of course very sensitive and subject to strict Census confidentiality procedures. As documented in Abowd, Haltiwanger, and Lane (2004), "only authorized researchers working from Census-controlled facilities have worked with the LEHD microdata; however, major efforts to make the data available to external researchers are underway." Since 2005, external researchers may access the LEHD data infra-

structure in the Census-National Science Foundation RDCs, and several research projects are already under way. Public-use summary data from the Quarterly Workforce Indicators are currently available, but synthetic data may facilitate the release of customized LEHD microdata products.

A.2.4 National Establishment Time Series

The National Establishment Time Series (NETS), another nongovernment data source, captures business relocations and employment change (job destruction and creation) for business establishments disaggregated at fine geographic detail. The data were developed from D&B data by Walls and Associates. As discussed earlier, the primary goal of D&B data is to sell information to businesses about businesses for decision-making purposes. Walls and Associates linked the data with the goal of constructing longitudinal files for studying business dynamics (Neumark et al., 2005a, p. 10).

Neumark et al. (2005b) used NETS to analyze employment growth in California; they also provide a detailed description of the data set. The authors were given access to annual data for the universe of business establishments located in California between 1989 and 2002 (there were about 3.5 million) for the purpose of examining the extent to which job creation and losses were attributable to interstate business relocation and to business births and expansions and contractions and deaths.[21] Regarding data quality, the authors conclude that NETS is a "reliable data source although not without limitations" (p. 3). They note, in particular, the rounding of employment figures and that short-term changes are not picked up particularly well (p. 31). They suggest that analysts should use three years of data to minimize the effects of these shortcomings. Davis et al. (1996) provide a discussion of limitations of D&B data more broadly. One advantage of these private data sources is that they are not subject to the access restrictions that handcuff the statistical agencies.

A.3 DATA SOURCES DESIGNED TO IMPROVE COVERAGE OF SMALL AND YOUNG BUSINESSES

A.3.1 Survey of Small Business Finances

In addition to the Census Bureau's BR, which has included data on nonemployers annually since 1994, there are a number of specialty surveys that focus on small business. A particularly important one is the Survey of

[21]They found that the latter is responsible for almost all employment change in California.

Small Business Finances (SSBF), which has been conducted by the Board of Governors of the Federal Reserve, with assistance from the National Opinion Research Center, in 1987, 1993, 1998, and 2003. (The Federal Reserve Board intends to discontinue the SSBF.) The SSBF is a nationally representative sample of small and minority-owned businesses with fewer than 500 employees screened for eligibility using D&B data files. Interviews were ultimately conducted with 3,500 firms from this class, including oversamples of African American-, Asian American-, and Hispanic American-owned firms. The Federal Reserve's objective with the SSBF was to collect information to better understand overall finances and credit conditions that small firms face, including:

- factors that affect prices and availability of credit to small businesses;
- effects that bank consolidation may have on the availability of credit and other financial services;
- characteristics of small businesses and how these characteristics influence their credit needs;
- experiences that small businesses have with credit applications;
- the impact that government regulations may have on small business credit access; and
- the financial and nonfinancial sources used by small businesses for their financing needs (http://www.federalreserve.gov/ssbf/).

The survey includes information on income and expenses, assets and liabilities, and financing sources, much of which is not available from any other sources. SSBF data are used to evaluate the impact of public policies, bank mergers and consolidations, and the rise in interstate banking on small businesses of different sizes, locations, and ownership characteristics (http://www.norc.uchicago.edu/studies/economic.htm). Summary analyses of the data are published in the "Report to Congress on the Availability of Credit to Small Business" that is produced by the Federal Reserve every five years. In addition, the SSBF has provided the foundation for a wide variety of academic research—for example, analyses of shifts in lending from the banking to the nonbanking sectors, bank mergers and consolidations, and the rise in interstate banking, as each relates to small businesses.

A.3.2 Small Business Administration-Funded Data Sources

The U.S. Small Business Administration (SBA) is the government agency most directly concerned with small business interests.[22] Its mission is to

[22]See Armington (2004) for a history of the SBA's role in the development of data on small businesses.

"maintain and strengthen the nation's economy by aiding, counseling, assisting and protecting the interests of small businesses and by helping families and businesses recover from national disasters" (www.sba.gov). Since Congress instructed the SBA in 1979 to begin developing data sets covering small businesses, the agency has worked to push forward data collection for use in studying firm dynamics—particularly job creation, attributable to smaller businesses. Generally the SBA is interested in the enterprise unit of analysis (as opposed to small establishments that are owned or controlled by large firms).

The SBA is involved in a number of programs; one of particular note is the Business Information Tracking Series (BITS), another effort to edit and longitudinally link archived data. BITS, which at times has also been known as the Longitudinal Establishment and Enterprise Microdata (LEEM), is partially funded by the SBA and carried out by the Census Bureau using SSEL-based data. Essentially, the SBA creates an enterprise version of County Business Patterns, called the Statistics of U.S. Businesses (SUSB), that consists of annual observations on each business and includes data on number of firms, number of establishments, employment, and annual payroll of firms categorized by location and industry (http://www.sba.gov/advo/research/data.html). SUSB provides annual snapshots on businesses from 1988 to 2002; the BITS program linked these records annually from 1989 to 2001. BITS provides data on private-sector establishments (single physical locations) with positive payroll; it includes, for each year, employment, annual payroll, 4-digit Standard Industrial Classifications, location, start year, legal entity, and total employment. BITS is structured to identify firm births and deaths, expansions and contractions, and mergers and acquisitions and to examine job flows. Firms are tracked using identifiers designed to remain unchanged even if there is a change in legal or ownership status. Among its limitations, BITS does not include the self-employed, is characterized by a long lag in production, and tracks only establishments (not firms).

The SBA publishes aggregate statistics on numbers of business formations and the distribution and growth of large versus small businesses over time from these data; however, given its business list foundations, microdata are available only to researchers who successfully apply to use them at Census RDCs.

A.3.3 Kauffman Firm Survey

The Kauffman Firm Survey is a new initiative designed to produce "a data set of publicly accessible research on new businesses and their development in the United States." The Kauffman Foundation began working with

Mathematica Policy Research, Inc., to develop and administer this survey in 2005. The Kauffman Firm Survey is a longitudinal survey of the principals of 5,000 firms, sampled from D&B, that started operations in 2004. The survey is oriented primarily to generate data on the financial development of new businesses in their first four years of existence. Surveys will be conducted by either telephone or on the Internet, and owners of these businesses will be asked about the characteristics of the business, about the financing of business operations, and about characteristics of the owner(s). Three follow-up interviews are planned with these businesses in 2006-2008 (http://www.mathematica-mpr.com/surveys/kauffmanfirm.asp). Data from this survey are not currently available, but they will ultimately provide publicly available longitudinal data on new firms (http://www.kauffman.org/research.cfm).

A.4 EMPLOYMENT STATISTICS

Two monthly surveys underpin measurement of employment levels and trends, over time—the Current Population Survey (CPS), a household survey, and the CES survey, a payroll or establishment survey. Employment estimates from each are published monthly. In addition to these two surveys, the BLS houses the Job Openings and Labor Turnover (JOLTS) program, which produces monthly data on job openings, new hires, and voluntary and involuntary separations.

The CPS, conducted by the Census Bureau and BLS, has been in existence for more than 50 years. It is based on a survey of approximately 60,000 households designed to estimate total employment of persons age 16 and over in the civilian noninstitutional population. The CPS captures employment broadly, including unincorporated self-employed, unpaid family workers, and agricultural workers; it also collects demographic and supplemental information.

The CES is a monthly survey of 160,000 businesses and government agencies covering approximately 400,000 establishments. It is a simple random sample stratified by state, industry, and size, which produces estimates of the number of nonfarm payroll jobs, hours, and earnings estimates based on payroll records of business establishments. CES counts jobs, meaning that multiple jobholders are overrepresented (from an employment perspective) and the self-employed are excluded—which points to the importance of household data for the production of comprehensive employment statistics. QCEW micro files serve as the sampling frame for the CES; the quarterly LBD is used to identify new business births and deaths.

A.4.1 CPS

The CPS is BLS's most widely used household survey. It is used in the production of monthly estimates of unemployed persons in the United States, providing information on employment along occupational, industry, and other dimensions. The CPS is a unique source of "business data" in that the households, rather than businesses, serve as the sampling unit. Microdata going back to 1962 are publicly available.

Specifications and uses of the CPS are documented in great detail at the BLS web site (http://www.bls.census.gov/cps/cpsmain.htm) and elsewhere. The survey, conducted by Census for BLS, involves a monthly sample of 60,000 in the civilian population, age 15 and older. The survey asks respondents for basic demographic information—age, gender, race, marital status, education, immigrant status, family structure, and labor market status (for those age 16 and over)—in addition to the employment status questions. Respondents are asked if they work for private firms or government or are self-employed, and they are asked to provide hours worked, occupation, industry, and earnings information.

The CPS is valuable for identifying new businesses, detected when an individual goes from employed to self-employed status from one period to another. Supplements provide detailed information on veterans, computer/Internet use at home and at work, adult education, health care, and pension coverage. The CPS is structured as a panel, in which each housing unit is sampled for four consecutive months, out for eight, and back in for four. This creates some opportunities for studying self-employed and entrepreneurship.

A.4.2 CES

The QCEW, including the payroll records it receives from the state employment security agencies (SESAs), is a benchmark data source for the CES. The CES provides employment, hours, and earnings estimates based on the SESA payroll records. Most of the CES employment series begin in 1990 for the reconstructed NAICS system, although employment by industry supersector is available since 1939. The CES is based on a sample of about 400,000 business establishments (160,000 firms) in over 1,150 industries (hours and earnings data are available for about 850 industries). The LBD, stratified by state, industry, and employment size, serves as the sampling frame. Geographic location of establishments is designated using Census-defined MSA guidelines. Beginning with the May 2003 data, the CES began publishing data organized using the NAICS classification system (http://www.bls.gov/ces/home.htm).

Series for employment hours and earnings at detailed industry and

geographic levels are collected and published monthly. In addition to the online database, the CES program produces a monthly news release and the monthly periodical *Employment & Earnings*, which are released on a three- and five-week lag, respectively. The database includes information on total employment, number of women employed, average weekly and monthly hours and earnings, number of production or nonsupervisory workers, and average weekly overtime hours in manufacturing industries. Microdata are not publicly available.

Like the BEL and QCEW sampling frames, the CES does not include nonemployers or detailed characteristics of business owners (http://www.bls.gov/ces/cesprog.htm).[23] The hourly employment data are used in the BLS industry productivity measures. However, the CES program collects hours and earning data only for production workers (primarily from the natural resources and mining and manufacturing sectors) and nonsupervisory workers (only from the service-providing industries). Employment data are collected for all workers, including production and nonsupervisory workers. BLS is currently collecting hours and earning data for all employees and plans to publish these data in early 2007 (http://www.bls.gov/ces/cesww.htm). Triplett and Bosworth (2004) argue that "with the huge change in workplace organization and management in recent years, the boundary between 'production' and 'non-production' workers has lost its meaning," and that the same applies to supervisory and nonsupervisory workers outside manufacturing. They conclude that there is a clear need for BLS to collect information on hours of work for all workers, as well as information on changes in labor quality at the industry level.

A.4.3 Job Openings and Labor Turnover

The BLS JOLTS program provides monthly data, based on a sample of about 16,000 establishments nationwide, on job openings, new hires, and both voluntary and involuntary separations. The definition of a job opening or vacancy requires that a specific position exists, that work could start

[23]BLS's annual national compensation survey—a redesign of the occupational compensation survey, the employment cost index, and the employee benefit survey—provides additional details on employee compensation costs, occupational earnings, and a range of worker benefits. Other agencies collect data on employee benefits as well. The Centers for Disease Control and Prevention conduct the National Employer Health Insurance Survey, which provides information on employer-sponsored health insurance; the Agency for Healthcare Research and Quality conducts the Medical Expenditure Panel Survey, providing information on health insurance coverage by firms of various sizes and an overview of access to health insurance and care.

within 30 days, and that the employer is actively recruiting from outside the establishment with the opening. Data are available in a national series and for four regions. Because the JOLTS data series has only recently been developed—data are available back to December 2000—it is still rather limited for business cycle analyses. However, once a sufficiently long time series has been established, the JOLTS series on hiring activity and job turnover will become useful for a range of labor market analyses not currently possible with other data sources.

JOLTS has already been used for assessing the reliability of the Conference Board's help-wanted index as a measure of job vacancies, which for years has been the primary monthly indicator of labor demand (http://www.bls.gov/jlt/). Since it has been available over a long time period and at the level of disaggregated regions, the help-wanted advertising index has been the leading data source for assessing important aspects of the labor market operations, "such as the effectiveness of the job-matching process. Depending on the intended application, the help-wanted series may be most useful when adjusted for underlying trends that are unrelated to labor market conditions" (http://www.frbsf.org/publications/economics/letter/2005/el2005-02.html).

A.5 DATA ON THE SELF-EMPLOYED, ENTREPRENEURS, AND BUSINESS GESTATION

The vast majority of information on businesses is collected from employer firms; as a result, less is known about the self-employed. One exception among the data sources described so far is the CPS. In this section, we describe data sources that are useful for research on emerging businesses, the self-employed, and entrepreneurs; in the process, we also discuss some of the topics that can be illuminated with data collection efforts centered on households.

A.5.1 Household Surveys: CPS and the Kauffman Index of Entrepreneurial Activity

Research by Robert Fairlie, who reported to the panel at its September 2005 meeting, articulated the value of household data for research related to small business dynamics. Fairlie has used CPS data to analyze the self-employed working in their own incorporated or unincorporated businesses (BLS published estimates do not include incorporated business owners), capturing both employer and nonemployer businesses. The CPS is designed primarily to collect employment information, but, since it also captures demographic and other supplemental information, a richer picture can be created.

The four months in, eight months out, four months in structure of the CPS makes it possible to match individuals from the first survey period to the next, as well as month to month to create panel data. Fairlie has been able to achieve annual match rates of about 70 percent. By linking CPS files over time, longitudinal data can be created, a situation that allows for the examination of business creations. Entrepreneurship rates for the 1980-2001 period, calculated using this method, are presented in Fairlie (2004). If individuals are tracked over an extended period of time, up to 3 to 4 years, the results of their efforts in operating a business could be tracked when the business files a Schedule C or, if it hires employees, in the state Unemployment Insurance and federal Social Security establishment registries. Entrepreneurs are defined as individuals who report owning a business who did not own a business during the previous survey year. A complication with this method arises if the time spent starting a business is not considered new business activity without any information on economic activity of the business entity itself.

Other data sources have been used to estimate self-employment among population subgroups. For example, BLS's National Longitudinal Survey of Youth (NLSY), a nationally representative sample of 12,686 young men and women who were between the ages of 14 to 22 in 1979, has been used to generate rates for economically disadvantaged groups. Individuals in the NLSY were interviewed annually through 1994 and are currently interviewed biennially. The NLSY collects information through an event history format and focuses primarily on labor force behavior, although it collects information on a wide range of variables, including child care costs, welfare receipts, aptitude measures, and school achievement. A second cohort of 9,000 youths were surveyed in the NLSY beginning in 1997 and are interviewed on an annual basis.

Similar work has been done to create the Kauffman Index of Entrepreneurial Activity, which uses matched data from the 1996-2004 CPS to develop a new measure of entrepreneurship. All individuals ages 20-64 who do not own a business as their main job are identified in the first survey month. By matching CPS files, it is then determined whether these individuals own a business as their main job with 15 or more usual hours worked per week in the following survey month. The Kauffman index is thus defined as the percentage of the population of nonbusiness owners who start a business each month. An average of 0.36 percent of the working-age population created a business each month, representing approximately 550,000 new businesses per month.[24]

[24]Based on presentation by Robert Fairlie, University of California, Santa Cruz, to the Panel on Measuring Business Formation, Dynamics, and Performance, September 23, 2005, National Research Council, Washington, DC

The advantages of household data include comparatively large sample sizes and long time series and more timely estimates of business ownership and entrepreneurship. Some of the disadvantages involve limited information on business outcomes, employment, and customers. These relative benefits reveal the attraction of integrating household and business data, which creates the potential to pull the best information from both.

A.5.2 Survey of Business Owners (SBO)

The Census Bureau's SBO generates information on the characteristics of business owners (e.g., gender, race) and their sources of financing. The richness of the data is enhanced because they can be linked with other longitudinal survey and administrative data collected by the Census Bureau. The SBO collects information on the minority status of owners and allows for complex responses in which each owner can report a self-identified race, as well as multiple racial groups. The sampling frame includes "all firms operating during 2002 with receipts of $1,000 or more that filed tax forms as individual proprietorships, partnerships, and any type of corporation, except those classified as: agricultural production; domestically scheduled airlines; railroads; U.S. Postal Service; mutual funds (except real estate investment trusts); religious grant operations; private households and religious organizations; public administration; and government" (www. census.gov/Press-Release/www/2005/sbo2002_ presentation.ppt). The IRS compiles for the Census Bureau a list of all companies filing any of the following forms: IRS Form 1040, Schedule C (for an individual proprietorship or a self-employed person); 1065 (for a partnership); any one of the 1120 corporation tax forms; and 941 (Employer's Quarterly Federal Tax Return) (http://www.census.gov/econ/census02/text/sbo/sbomethodology. htm).

Separate SBO reports for the Census Bureau's Surveys of Minority and Women Owned Business Enterprises include more detailed data by geographic area, kind of business, and size of business (full details on data characteristics and reliability can be found at http://www.census.gov/csd/sbo/intro2002SBO.htm). Conducted in conjunction with the quinquennial economic census program, these surveys are used in constructing and publishing the Surveys of Minority-Owned and Women-Owned Business Enterprises Company Statistics Series. Number of firms, sales and receipts, paid employees, and annual payroll are included in the published data, which are presented by geographic area, industry, firm size, and legal form of organization. Microdata are protected under Title 13 confidentiality guidelines and are available for statistical purposes to researchers with special sworn status with the Census Bureau (http://www.census.gov/csd/mwb/).

A.5.3 Panel Study on Entrepreneurial Dynamics (PSED)

Organizations outside government have developed keen interest in measuring entrepreneurial activities. The PSED, sponsored by the Kauffman Foundation, is one such manifestation of this interest. The PSED is a nationally representative database for the United States designed to enhance understanding of the business start-up phenomenon. The PSED is a longitudinal sample of over 64,000 U.S. households that were contacted to find individuals who were actively engaged in starting new businesses. Resulting data are intended to promote research into the business gestation process—the period before the business actually produces output. The PSED includes information on the proportion and characteristics of the adult population involved in attempts to start new businesses, the kinds of activities nascent entrepreneurs undertake during the business start-up process, and the proportion and characteristics of the start-up efforts that become infant firms. Prevalence rates for nascent entrepreneurs are reported by gender and ethnicity (white, black, and Hispanic) on such demographic variables as age, education, household income, and urban context (Gartner et al., 2004; Reynolds, 2006).

The PSED focuses on four fundamental questions:

- Who is involved in starting businesses in the United States?
- How do they go about the process of starting companies?
- Which of these business start-up efforts are likely to result in new firms?
- Why are some of these business start-up efforts successful in creating high-growth businesses? (http://research.kauffman.org/cwp/app manager/ research/)

Four panels of data are currently available covering the time period 1998 to 2003. Additional work has begun that will follow a cohort of "nascent entrepreneurs" for three years. Data from the first-year interviews are scheduled to be available in 2006. The data are being used to address a number of research questions—for example, to estimate the number of individuals or teams of individuals in the United States attempting to start a business at any given time, the variation in start-up rates among minority groups, and the impact of education on entrepreneurship. Data from the PSED are maintained and made available for download by the University of Michigan's Institute for Social Research (http://www.psed.isr.umich.edu).

A.5.4 Global Entrepreneurship Monitor (GEM)

A recent Organisation for Economic Co-operation and Development (OECD) report (Vale, 2006) provides an inventory of the different sources of data on start-ups in the OECD member countries. One data source described in the study is the GEM; in this project, a new sample is drawn each year from adult populations to generate a harmonized set of cross-national comparisons. The GEM collects data on various aspects of entrepreneurship through a series of coordinated household surveys in a gradually increasing number of countries worldwide. The GEM has focused on the study of early-stage entrepreneurial activity and is moving into "analyses of the existence and characteristics of established business owners; the degree of innovativeness, competitiveness, and growth expectations of early-stage and established business owners; and the existence and characteristics of social environments conducive to entrepreneurship" (Global Entrepreneurship Monitor, 2004; http://www.gemconsortium.org/).

Using the conceptual framework and methodology developed for the initial PSED, the GEM program was developed in 1998 to provide harmonized cross-national comparisons of the prevalence of adults participating in new firm creation. This was accomplished by commissioning surveys of the adult population in participating countries using a common interview schedule and consolidating and standardizing the responses; the result is the capacity to compare the proportion of the adult population (ages 18-64) in each country that is actively engaged in new firm creation. The major comparisons utilize the Total Entrepreneurial Activity Index, reflecting both those in the start-up process as well as owner-managers of new firms up to 42 months old. These interviews were supplemented by personal interviews with 30-80 national experts in entrepreneurship in each country, standardized expert questionnaire responses, and an assembly of a considerable amount of harmonized national data from standardized sources (Reynolds et al., 2005).

The GEM also produces a report on women and entrepreneurship based on survey data from more than 107,000 respondents in 35 countries. The report attempts to quantitatively characterize and describe patterns of behavior among women entrepreneurs relative to those of men and to measure the gender gap for entrepreneurial activity internationally. The first annual report was produced on 10 countries in 1999. By 2005, over 44 countries, as well as specialized regions in some countries, had been involved for one or more years. An annual report summarizes the major cross-national findings supplemented by special topic reports and annual country reports. Over 100 of these reports and the major data sets, with a three-year lag, are provided on the project web site (www. gemconsortium. org).

A.5.5 American Time Use Survey (ATUS)

Other federal data collections exist that, although not specifically designed to measure entrepreneurial activity, may eventually be useful. One example is BLS's recently launched ATUS. The ATUS includes data on time spent working, sleeping, shopping, volunteering, participating in leisure activities, etc. It also includes self-employment identifiers and data on hours worked. In addition to its time diary, the ATUS collects demographic and labor force information, as well as summary questions on child care activities, paid work activities, and absences from home (Frazis and Stewart, 2004).

The ATUS diary can be used to assess the validity of claims that household surveys overstate hours worked (Robinson and Bostrom, 1994). A comparison of data on hours worked from the CPS and the ATUS (2003 data) provides some support for this claim. The average number of weekly hours worked estimated from the ATUS fell in the 37.3 to 37.9 range—depending on which work-related activities are included—relative to about 39 hours reported in the CPS (Frazis and Stewart, 2004). One possible explanation is that hours worked during CPS reference weeks (the week including the 12th of the month) are higher than nonreference weeks. In fact, in their comparison work, Frazis and Stewart found that "estimates of actual hours worked from the CPS are very close to (ATUS) time-diary estimates for the CPS reference week." In other research, Song (2005) and Eldridge and Pabilonia (2005) used ATUS data to investigate the incidence of people working at home and its relation to pay status and length of hours worked. ATUS data can also be used to compare how self-employed workers spend their time relative to wage and salary workers. However, "work" itself is a black box category, and not much can currently be done to measure specific entrepreneurial activities.

A.5.6 The American Community Survey (ACS)

The Census Bureau's ACS will generate household-based data that will also be useful for studying self-employment trends. Question 35 of the survey asks respondents about their current or most recent job activity—specifically, whether the person was "an employee of a private for profit company; an employee of a private not for profit, tax-exempt, or charitable organization; a local government employee; a state government employee; a federal government employee; self-employed in own not incorporated business, professional practice, or farm; self-employed in own incorporated business, professional practice or farm; working without pay in a family business or farm." Question 41 asks for "self-employment income from own non-farm business or farm business, including

proprietorships and partnerships (report net income after business expenses)" http://www.census.gov/acs/www/SBasics/SQuest/SQuest1. htm.

The ACS data undergo extensive computer editing to correct reporting deficiencies and improve consistency of the income reports. For example, "if people reported they were self employed on their own farm, not incorporated, but had reported only wage and salary earnings, the latter amount was shifted to self-employment income."[25] The ACS data are limited in that income is often underreported. In addition, the earnings data generated are not directly comparable with those from the SSA records, since SSA data are based on employer reports and income tax returns for the self-employed. Furthermore, SSA excludes some civilian government and nonprofit organization employees, workers covered by the Railroad Retirement Act, and people whose earnings are insufficient for the Social Security program.

A.6 DATA COVERAGE OF SPECIAL SECTORS: AGRICULTURE, NONPROFITS, AND E-COMMERCE

While this report does not deal with or make recommendations specifically about issues that are unique to the agricultural sector, we point out here that there are numerous data sources covering the sector. In fact, relative to its quantitative role in the economy, more is spent in the production of statistics on agricultural output (and other variables) than on other areas of the economy, including those that are growing rapidly.

A.6.1 Agricultural Resource Management Survey (ARMS)[26]

The Economic Research Service (ERS) and the National Agricultural Statistical Service of the U.S. Department of Agriculture (USDA) are joint sponsors of the annual ARMS, the primary source of information on "the financial condition (including debt levels), production practices, resource use, and economic well-being of America's farm households." The ARMS provides data on the business side of farms (i.e., field-level farm practices and economics of the business) as well as household characteristics, such as age, education, farm and off-farm work and income, and family living expenses, making ARMS the most broad-based source of national agricultural business data.

[25]American Community Survey: 2004 Subject Definitions (www.hawaii.gov/dbedt/info/census/acs/acs_subject_definitions_8_05.pdf).

[26]Information on this survey can be found at the USDA web site: http://www.ers.usda.gov/data/arms/GlobalAbout.htm#Use.

The ARMS data provide the basis for various USDA estimates, including the annual cost-of-production estimates the department is congressionally mandated to produce for over 15 commodities and the annual estimates of average and net farm income, which in turn are used by the Bureau of Economic Analysis (BEA) to develop GDP and personal income estimates. In addition, the Food and Agriculture Act of 1977 requires USDA to produce the Annual Report on Family Farms. In preparing this report, ERS draws on the ARMS data for information on a host of relationships, including:

- farm participation in agricultural programs and the distribution of farm program payments;
- the structure and organization of farms, including family and nonfamily ownership;
- the use of new production technologies and other management practices;
- farm use of credit;
- farmers' participation in off-farm employment; and
- identifying the characteristics of producers purchasing crop insurance.

A.6.2 Nonprofit Organizations[27]

While there are gaps in data for studying the performance and economic contribution of nonprofit organizations, many of the data that are collected are publicly available because confidentiality constraints on information (including financial data) are very different. The IRS maintains a continuously updated registry of tax-exempt nonprofit organizations, which, in turn, is incorporated into the IRS Business Master File. Unlike other components of the master file, the IRS is allowed to provide public access to information on Forms 900 filed by nonprofits, a condition for tax-exempt status. Information becomes available to the public roughly six months after the IRS rules that an organization qualifies for tax-exempt status. The IRS registry of exempt organizations can be used to identify births and deaths in the sector. Registry data have been used to measure elapsed time from birth to filing Form 990 (David, Pollak, and Arnsberger, 2005) and, with a longer lag, also transitions to inactive status.

[27]To a large extent this description is a summary of a presentation to the panel by Martin David, Urban Institute, on the state of data for measuring activity in the nonprofit sector.

Census Bureau lists include Business Master File and Form 941 information on nonprofit organizations, as does the QCEW at BLS. The QCEW can be used to measure employment growth for new organizations; however, since nonprofit status is not identified, estimates (such as Knaup, 2005) cannot easily be broken out by for-profit and nonprofit status. However, researchers from the Johns Hopkins Center for Civic Society Studies have used QCEW data via the BLS outside researcher program to develop a way to identify nonprofit organizations in the QCEW and produce employment statistics on the sector. The joint initiative method involves identifying tax-exempt firms in the data sets by "matching employer identification numbers on the QCEW files with those on the exempt organization master file, maintained by the Internal Revenue Service." The Form 990 information from the IRS is limited in that the data are organization-based, rather than establishment-level based (Salamon and Sokolowski, 2005). Unlike the Census Bureau's economic census, the QCEW has broader coverage of the nonprofit sector and is more timely. The decennial population census and the CPS cover the nonprofit sector; however, research shows that the self-identification of the profit or nonprofit status of a workplace is questionable.

There is high demand for small-area statistics on the numbers of nonprofit organizations and employees. In principle, such estimates could be produced, as often as quarterly, from states and metropolitan areas from Form 941 or QCEW data. Employment information from QCEW data are more precise since organizations are disaggregated to the establishment level. Also, employment information is not collected from organizations with less than $100,000 annual revenue that file Form 990-EZ. Organizations with less than $25,000 annual revenues over three years do not need to file at all, and there are not much data on slightly larger organizations without paid employees.

These sources say little about the use of volunteer labor, an important input for nonprofit operation. Estimates of the supply of volunteer labor can be produced by various household data collections, particularly the CPS supplement and the newly developed ATUS. Use of this labor, particularly at the establishment level, is still largely missing.

Martin David suggested to the panel that microdata that include the exempt status of the organization should be created and made available to researchers at the Census and BLS data centers with restricted access. He noted further that the NAICS industrial classification is insufficient for nonprofit organizations. The National Taxonomy of Exempt Entities (NTEE), which is used by the IRS to classify annual information returns, is more detailed (for example, it distinguishes between elementary and high schools, which NAICS does not). Carrying over NTEE from the source

documents at the organization level to statistical coverage at the establishment level would be useful. His other suggestions were to require that Form 990 be filed more promptly after the end of the fiscal year and to require it to identify revenue from government contracts to improve information for measuring balance of private financing versus government financing of nonprofits.

The Urban Institute's National Center for Charitable Statistics has created a digital census of charitable nonprofit organizations (classified as IRS 501(c)(3) entities), built from Form 990 data, that contains most of the fields from these forms. Industry coding is also incorporated, and the national taxonomy of exempt entities used in creating the digital census improves classification of organizations, which creates value added to the IRS registry. Microdata are linked into a panel structure using EINs. New data are released annually and are accessible to researchers.

A.6.3 E-Commerce[28]

The impact of electronic commerce (e-commerce) on the U.S. economy is growing and is now widely discussed; however, there is a significant data gap in the official, national economic statistics. The Census Bureau and BEA are involved in efforts to fill this data gap by measuring aspects of the "new economy." Much of the initial efforts were oriented toward simply defining e-commerce and digital and electronic economic activities (Tehan, 2003). A number of basic questions emerged:

- How is business-to-business and business-to-consumer e-commerce impacting the accuracy of labor surveys?
- What are the goods and services choices, characteristics, and prices offered?
- How difficult is it to track international transactions as well as business costs and productivity (Tehan, 2003)?

Gaps in data on e-commerce (e.g., the extent to which computers and the Internet have reduced entry barriers for small businesses, making them better equipped to compete in the economy) have kept researchers from fully exploring questions involving the extent to which new technologies are altering business dynamics.

The taxonomy of e-commerce continues to evolve to reflect that clear distinctions are not always possible—online retailers expand their inven-

[28]Much of this information is taken from Berney (1999) and from the *Report for Congress, E-Commerce Statistics: Explanation and Sources* (updated June 4, 2003).

tory, and brick-and-mortar businesses develop online components. The Census Bureau collected e-commerce data in four surveys, to measure various aspects of economic activity, including shipments for manufacturing, wholesale and retail trade sales, and service industry revenues. To capture whether firms were conducting sales online and the volume of these sales, the Census Bureau added two questions to its monthly retail trade survey in fall 1999. In addition, the Census Bureau developed the U.S. Department of Commerce E-Stats web site devoted to "measuring the electronic economy," which covers NAICS industries, or 70 percent of economic activity from the 1997 economic census (Tehan, 2003, pp. 3-5). BEA has considered whether e-commerce can be accurately measured and has proposed "a comprehensive measure of e-business and high-tech that would measure the new economy in a comprehensive and consistent fashion" (Tehan, 2003, p. 6).

At the Understanding the Digital Economy conference in May 1999, John Haltiwanger noted that existing databases were unlikely to have the capacity in the near term to measure what is happening in the digital economy, particularly by firm size. The LEEM file, discussed above, can be used to determine growth rates, the geography of fast-growing firms, birth and death information, and mergers and acquisitions. The data, which existed in the LEEM file, were not used to advance understanding of the digital economy, because "the problems were that the digital economy would have to have been defined by the 1987 SIC code designations which are outdated and would have covered only the time period 1988-1995" (Berney, 1999, p. 3). In addition, Berney speculated that the NAICS codes will also have difficulty keeping up with this continuously evolving segment of the economy.

The Statistics Canada model provides a way to both reduce the bureaucratic lags from the cooperation of multiple data collection agencies and protect the individual agency budgets. "Until the federal statistical agencies can make more current data available, researchers will need to rely on survey information by private firms and trade associations to analyze what is going on in e-commerce and the digital economy" (Berney, 1999, p. 3). Microsoft, Nathan Associates, the International Data Corporation, and the Yankee Group have been involved in developing databases on the digital economy and e-commerce. The *2003 Report to Congress* lists web addresses for e-commerce statistics as of June 2003, including government, academic, and private research firms.

A.7 FINANCIAL DATA

When measuring and evaluating entrepreneurial activity, it is important to identify financial sources and to differentiate between venture-backed companies and others. Current data on financing and investments

are insufficient in terms of analyzing how financing interacts with the evolution of new firms. In his presentation to the panel, Steven Kaplan (University of Chicago Graduate School of Business) noted the need to integrate financial data with performance data. Given existing databases, it is, for example, very difficult to study the extent to which economic activity and job growth are driven by venture capital.

Existing databases include Venture Economics, Venture One, the Statistics of Income (SOI), Compustat, the Annual Capital Expenditures Survey (ACES), and the Kauffman Financial and Business Database (KFBD). Currently, the databases are not linked to one another. Venture Economics, owned by Thompson, is useful for valuation history. The Dow Jones Venture One includes higher quality, clean data and more valuation data and data on people than Venture Economics. However, the data from Venture One are difficult to access, compared with Venture Economics data, which are accessible for a fee.

SOI corporate data are the only publicly available source of corporate financial information; data products for S-corporations—those with 75 or fewer shareholders—are also available. The data, based on a stratified random sample of 130,000 preaudited returns, contain income statements, balance sheets and tax information, industry classification, identification of accounting periods and sizes of assets, receipts, and income taxes after credits. The SOI provides data annually to BEA on partnerships, as well as producing annual information on nonfarm sole proprietorships from Schedule C data.

Standard and Poor's produces Computstat, a database for all publicly traded firms in the U.S. stock market. It is geared mainly toward investors and attempts to standardize financial statements and accounting statement information on companies around the world. Compustat has been longitudinal since 1980 and includes such information as quarterly and annual income statements, balance sheets, and cash flow statements. Reporting units are identified by reporting company and a 4-digit SIC code and are business or industry segments, defined by the Financial Accounting Standards Board (FASB) Statement of Financial Accounting Standards No. 14 as: a component of an enterprise engaged in providing a product or service or a group of related products or services primarily to unaffiliated customers (i.e., customers outside of the enterprise) for profit (FASB Statement 14, p. 7).

In its words, the Census Bureau began the ACES as part of a comprehensive program designed to provide detailed and timely information on capital investment in new and used structures and equipment by nonfarm businesses. The survey sample includes approximately 46,000 employer companies and approximately 15,000 nonemployer companies. The survey prior to 1999 published capital expenditures data only for "97 industries

comprised of two-digit and selected three-digit industries from the Standard Industrial Classification (SIC) system. . . . Beginning with the 1999 ACES, for companies with employees, capital expenditures data are published for industries comprised primarily of three-digit and selected four-digit industries from the North American Industry Classification System (NAICS)" (http://www.census.gov/csd/ace/).

The Kauffman Foundation purchases data semiannually from D&B and uses them to develop the KFBD, a longitudinal file of annual data since 1996. More than 1 million records make up the KFBD, housing financial information on more than 300,000 firms and current demographic data for over 900,000 companies, each record including an annual balance sheet, an annual income statement, fourteen standard financial ratios, and various firm-level demographic items. The KFBD contains complete, consecutive financial statements for approximately 50,000 companies for a period of 3 years in length. These data may be sorted by industry (either NAICS or SIC codes), year started, number of employees, annual sales, minority ownership, and detailed location information, among other variables (http://research.kauffman.org/cwp/appmanager/research/researchDesktop).

REFERENCES

Abowd, J.M., J. Haltiwanger, and J. Lane
2004 Integrated longitudinal employee-employer data for the United States. *American Economic Review Papers and Proceedings* 94(May).

Armington, C.
2004 *Development of Business Data: Tracking Firm Counts, Growth, and Turnover by Size of Firms*. (Prepared for the SBA Office of Advocacy). Washington, DC: U.S. Small Business Administration.

Bartelsman, E.J., and M. Doms
2000 Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature* 38(3):569-594.

Berney, R.
1999 Comments. Prepared for the Small Business Session, Understanding the Digital Economy Conference, May 25-26, Department of Commerce, Washington, DC. Available: http://www.technology.gov/digeconomy/powerpoint/berney/index.html.

Boden, R., and A. Nucci
2004 Longitudinal Linking of Nonemployer Data to Nonemployer Data and the Business Register Over Time: Some Preliminary Findings. Paper presented at the Eastern Economic Association Annual Meeting, February, Washington, DC.

Clayton, R.L., A. Sadeghi, and D.M. Talan
2005 New Data on Business Employment Dynamics. Presented at the Federal Committee on Statistical Methodology Research Conference, November 14-16, Washington, DC.

David, M., T. Pollak, and P. Arnsberger
2005 *Compliance with Information Reporting: Exempt Organizations*. Available: http://www.irs.gov/pub/irs-soi/05david.pdf.

Davis, S.J., and J. Haltiwanger
1990 Gross job creation and destruction: Microeconomic evidence and macroeconomic implications. *NBER Macroeconomics Annual* 123-168.
1992 Gross job creation, gross job destruction, and employment reallocation. *Quarterly Journal of Economics* (August):819-864.

Davis, S.J., J. Haltiwanger, R.S. Jarmin, C.J. Krizan, J. Miranda, and A. Nucci
2006 *Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes*. (Center for Economic Studies Working Paper #06-04). Washington, DC: U.S. Census Bureau.

Davis, S.J., J. Haltiwanger, and S. Schuh
1996 *Job Creation and Destruction*. Cambridge, MA: MIT Press.

Dunne, T., M. Roberts, and L. Samuelson
1989 Plant turnover and gross employment flows in the U.S. manufacturing sector. *Journal of Labor Economics* 7(1):48-71.

Eldridge, L.P., and S.W. Pabilonia
2005 *Are Those Who Bring Work Home Really Working Longer Hours? Implications for BLS Productivity Measures*. Washington, DC: U.S. Bureau of Labor Statistics. Available: http://www.oecd.org/dataoecd/36/55/37490637.pdf.

Fairlie, R.W.
2004 Recent trends in ethnic and racial self-employment. *Small Business Economics* 23(3):203-218.

Financial Accounting Standards Board
1976 Financial Reporting for Segments of a Business Enterprise. In *Statement of Financial Accounting Standards No. 14*. Norwalk, CT: Financial Accounting Foundation.

Frazis, H., and J. Stewart
2004 *Where Does the Time Go? Concepts and Measurement in the American Time Use Survey*. Washington, DC: U.S. Bureau of Labor Statistics. Available: http://www.nber.org/books/CRIW03-BH/frazis-stewart3-24-05.pdf.

Gartner, W.B., K.G. Shaver, N.M. Carter, and P.D. Reynolds, eds.
2004 *Handbook of Entrepreneurial Dynamics: The Process of Business Creation*. Thousand Oaks, CA: Sage Publications.

Getz, P., J. Kropf, R. Clayton, R.E. Detlefsen, P. Hanczaryk, R. Jarmin, and E.D. Walker
2005 *Births and Deaths in Business Surveys*. Presentation to the Federal Economic Statistics Advisory Committee, June 10, Washington, DC. Available: http://www.bea.gov/about/pdf/BirthsDeathsBusinessSurveys.pdf.

Global Entrepreneurship Monitor
2004 Global Entrepreneurship Monitor 2004 Global Report. Available: http://www.gemconsortium.org/document.asp?id=364.

Haviland, A., and B. Savych
2005 A Description and Analysis of Evolving Data Resources on Small Business. (RAND Institute for Social Justice Working Paper). Available: http://www.rand. org/pubs/working_papers/2005/RAND_WR293.pdf.

Jarmin, R., and J. Miranda
2002 *The Longitudinal Business Database*. (Center for Economic Studies, July). Washington, DC: U.S. Census Bureau. Available: http://webserver01.ces.census.gov/index. php/ces/1.00/cespapers/index.php/ces/1.00/cespapers?down_key=101647.

Knaup, A.E.
2005 Survival and longevity in the business employment dynamics data. *Monthly Labor Review* 128(5):50-56.

McGuckin, R.H., and S. Peck
1992 *Manufacturing Establishments Reclassified into New Industries: The Effect of the Survey Design Rules*. (Center for Economic Studies Discussion Paper #92-14). Washington, DC: U.S. Bureau of the Census.

Neumark, D., J. Zhang, and B. Wall
2005a *Employment Dynamics and Business Relocation: New Evidence from the National Establishment Time Series*. (NBER Working Paper #11647). Cambridge, MA: National Bureau of Economic Research.
2005b Are businesses fleeing the state? Interstate business relocation and employment change in California. *California Economic Policy* 1(4).

Okolie, C.
2004 Why size class methodology matters in analyses of net and gross job flows. *Monthly Labor Review* (July):3-12. Also available: http://www.bls.gov/opub/mlr/2004/07/art1full.pdf.

Pierce, K.
2005 Sole proprietorship returns, 2003. *Statistics of Income Bulletin* June 22. Washington, DC: Internal Revenue Service.

Pivetz, T.R., M.A. Searson, and J.R. Spletzer
2001 Measuring job flows and establishment flows with BLS longitudinal establishment microdata. *Monthly Labor Review* 124(40):13-20.

Reynolds, P.D.
2006 *New Firm Creation in the U.S.: A PSED I Overview*. Cheltenham, England: Edward Elgar.

Reynolds, P.D., N.S. Bosma, E. Autio, S. Hunt, N. De Bono, I. Servais, P. Lopez-Garcia, and N. Chin
2005 Global entrepreneurship monitor: Data collection design and implementation: 1998-2003. *Small Business Economics* (24):205-231.

Robinson, J.P., and A. Bostrom
1994 The overestimated workweek? What time diary measures suggest. *Monthly Labor Review* 117(8):11-23.

Salamon, L., and S.W. Sokolowski
2005 Nonprofit Organizations: New insights from QCEW data. *Monthly Labor Review* (September):19-26.

Salyers, E.
2004 Progress Report of the U.S. Census Bureau. Presented at the 18th Roundtable on Business Survey Frames. October 17-22, Beijing, China. Available: http://www.stats.gov.cn/english/18roundtable/papers/t20041230_402219768.htm.

Song,Y.
2005 *Working at Home: Pay Status and Timing*. Presented at the Annual Conference of the International Association for Time-Use Research, November 2-4, Halifax, Canada. Available: http://www1.union.edu/songy/WorkingatHome.pdf.

Spletzer, J.R., J. Faberman, A. Sadeghi, D.M. Talan, and R.L. Clayton
2004 Business employment dynamics: New data on gross job gains and losses. *Monthly Labor Review* (April):29-42.

Tehan, R.
2003 *E-Commerce statistics: Explanation and Sources*. (Congressional Research Service report). Damascus, MD: Penny Hill Press.

Triplett, J.E., and B.P. Bosworth
2004 *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Washington, DC: Brookings Institution Press.
Vale, S.
2006 *The International Comparability of Business Start-up Rates, Final Report*. Organisation for Economic Co-operation and Development, November, Paris, France. Available: http://www.olis.oecd.org.

TABLE A-1 BUSINESS DATA SETS[1]

The following acronyms are used in the table:

ATUS	American Time Use Survey
BEA	Bureau of Economic Analysis
BED	Business Employment Dynamics
BEL	Business Establishment List
BITS	Business Information Tracking Series
BLS	Bureau of Labor Statistics
BR	Business Register
CBP	County Business Patterns
CES	Current Employment Statistics
COS	Company Organization Survey
CPS	Current Population Survey
D&B	Dun and Bradstreet
DMI	Duns Market Identifiers
ECIS	Employment Cost Index Survey
EIN	Employer Identification Number
ES-202	Covered Employment and Wages (under the SESA UI program)
ETA	Employment and Training Administration
FRB	Federal Reserve Board
FTP	File Transfer Protocol
GDP	Gross Domestic Product
GEM	Global Entrepreneurship Monitor
ILBD	Integrated Longitudinal Business Database
IRS	Internal Revenue Service
KFBD	Kauffman Financial & Business Database
KFS	Kauffman Firm Survey
KIEA	Kauffman Index of Entrepreneurial Activity
LBD	Longitudinal Business Database
LDB	Longitudinal (establishment) Database
LED	Local Employment Dynamics
LEEM	Longitudinal Establishment and Enterprise Microdata
LEHD	Longitudinal Employer-Household Dynamics
LFO	Legal Form of Organization
LRD	Longitudinal Research Database
MSA	Metropolitan Statistical Area
NAICS	North American Industry Classification System

[1]Much of this information, and in some cases a good deal more, can be found on web pages of the statistical agencies and other data collection organizations.

NORC	National Opinion Research Center
NSF	National Science Foundation
OES	Occupational Employment Statistics
PPI	Producer Price Index
PSED	Panel Study of Entrepreneurial Dynamics
QCEW	Quarterly Census of Employment and Wages
QUI	Quarterly Unemployment Insurance
QWI	Quarterly Workforce Indicators
R&D	Research and Development
RDC	Research Data Center
SBO	Survey of Business Owners
SESA	State Employment Security Agency
SIC	Standard Industrial Classification
SMOBE	Survey of Minority-Owned Business Enterprises
SOI	Statistics of Income
S&P	Standard & Poor's
SSA	Social Security Administration
SSBF	Survey of Small Business Finances
SSEL	Standard Statistical Establishment List
SUSB	Statistics of U.S. Businesses
SWOBE	Surveys of Women-Owned Business Enterprises
UCFE	Unemployment Compensation for Federal Employees
UI	Unemployment Insurance

U.S. DEPARTMENT OF LABOR, BUREAU OF LABOR STATISTICS (BLS)

Quarterly Census of Employment and Wages (QCEW) [formerly the Business Establishment List (BEL)]

Purpose/uses	Serves as the sample frame for most BLS establishment surveys. Used to publish a quarterly count of employment and wages reported by employers (covers about 98% of U.S. jobs) available at the county, MSA, state and national levels, by industry. Used as benchmarking source for CES and OES. Used by BEA in calculating personal income components of GDP.
Design basics	Constructed mainly from UI system administrative records maintained by SESAs.
Frequency	Microdata continuously updated. County wage and employment data are published quarterly, 6 to 7 months after end of quarter.
Unit level	Establishment (EIN provides ability to aggregate to firm and industry levels and from county to national levels).
Coverage	Includes all establishments and workers covered by UI and UCFE programs. This amounts to about 8.4 million establishments—a near census—and 98% of all nonfarm wage and salary employment.
Content	Master file includes establishment names; full mailing and physical address; federal EIN; monthly employment; and quarterly wages by NAICS industry, county, ownership sector.
Limitations or lag time	Excludes nonemployer firms; no ownership characteristics; no government and only partial farm coverage; ownership links for multistate firms not comprehensive.
Accessibility of data	Microdata available only for statistical purposes and must be accessed onsite at the BLS research center in Washington, DC. Publicly available files include data on the number of establishments, monthly employment levels and quarterly wages, aggregated by NAICS industry, county, or ownership sector.

Business Employment Dynamics (BED)

Purpose/uses	Program produces a quarterly series of gross job gains and gross job losses statistics based on the universe of establishments covered in the QCEW. Used in CES to track establishment-level expansions, contractions, openings, and closings.
Design basics	Data are constructed using a multistep procedure to link QCEW microdata across periods; about 6.7 million private-sector employers are covered. Data become available about 8 months after the end of a given quarter.
Frequency	Quarterly, panel (1992 to present).
Unit level	Establishment (can be aggregated to firm level with EIN).
Coverage	All establishments subject to state UI laws and federal agencies subject to the UCFE program (about 98% of all employment).
Content	Monthly employment and wages; full address; and number and percent of gross jobs gained by opening and expanding establishments, gross jobs lost by closing and contracting establishments, establishments that are classified as openings, closings, expansions, and contractions.
Limitations or lag time	Excludes government employees, certain nonprofit employees, self-employed, private households, and establishments with zero employment. UI coverage differs by state and may change over time. EINs are imperfect for creating record linkages.
Accessibility of data	As with QCEW, public-use microdata are not available. Researchers must submit proposals and data can only be used at the BLS research center in Washington, DC.

Current Employment Statistics (CES)
[also known as Payroll Establishment Survey]

Purpose/uses	Provides employment, hours, and earnings estimates based on payroll records. Provides first economic indicator of current economic trends each month (with unemployment rate).
Design basics	Based on a sample of about 400,000 business establishments (160,000 firms). The LDB, stratified by state, industry, and employment size, serves as the sampling frame.
Frequency	Monthly
Unit level	Establishment
Coverage	Payroll employment for establishments in nonagricultural industries (over 1,150 industries). Hours and earnings data are collected from SESAs for about 850 industries.
Content	Total employment, full address, number of women employed, number of production or nonsupervisory workers, average hourly earnings, average weekly hours, average weekly earnings, and average weekly overtime hours in manufacturing industries.
Limitations or lag time	As with QCEW, there is no nonemployer, self-employed, or farm coverage, and no detailed owner or small firm characteristics. Geographic coding is available only by MSA. Establishments are not tracked over time and multiple jobholders are overrepresented.
Accessibility of data	Electronic access to selected indicator data is available. Microdata are not publicly available. Researcher can apply for access to the confidential microdata.

American Time Use Survey (ATUS)

Purpose/uses	Collects information on how people in the United States spend their time, including kinds of activities and time spent doing them. Used in preparation of BLS press releases and to produce categorical time use tables on ATUS web site.
Design basics	Sample frame is drawn from households that have completed their final month of interviews for the CPS, utilizing a stratified, 3-stage sample.
Frequency	Data have been collected since 2003, and they are published annually.
Unit level	Individual (household)
Coverage	Civilian noninstitutional population and workers ages 16 and over. For 2004 and 2005, approximately 27,000 cases yielded about 13,500 completed interviews; the survey was roughly 50% larger in 2003. Diaries are used to capture data spent on various activities.
Content	Data are collected on major activity categories (work, sleep, eating, etc.) and on selected variables such as earnings, school enrollment, selected demographics, household, labor force characteristics, and hours worked. There is also a self-employment identifier.
Limitations or lag time	Little information is collected on secondary activities (those done in combination with other activities) not collected. Estimates subject to nonsampling errors, particularly if nonresponse is correlated with time use.
Accessibility of data	Published tables and microdata files available on the ATUS web site.

BLS, continued

Job Openings and Labor Turnover Survey (JOLTS)

Purpose/uses	Data serve as demand-side indicators of labor shortages at the national level. Availability of unfilled jobs—the job openings rate—is an important measure of job market dynamics that complements measures of unemployment.
Design basics	Data from a sample of approximately 16,000 U.S. business establishments are collected on a voluntary basis. The sample frame consists of approximately 8 million establishments on the BLS' ES-202 QCEW file. Reference periods for total employment is the pay period that includes the 12th of the month; for job openings, it is the last business day of the month; for hires and separations, it is the entire calendar month.
Frequency	Data tables are released monthly.
Unit level	Establishment
Coverage	The survey covers all nonagricultural industries in the public and private sectors for the 50 states and the District of Columbia.
Content	Total employment, job openings, hires, quits, layoffs and discharges, and other separations.
Limitations or lag time	Data available only on a national level. No turnover rates by occupation.
Accessibility of data	Data are disseminated in a news release and through updated tables on the BLS website.

BLS AND U.S. CENSUS BUREAU

Current Population Survey (CPS)

Purpose/uses	Provides information on the labor force characteristics of the U.S. population. Data are used to calculate total employment (by occupation) and unemployment statistics. Used as sample frame for ATUS. Used to produce supplements on displaced workers, job tenure and occupational mobility. CPS data have also been used to construct the KIEA (1996 to 2004), a measure of business creation defined as the percentage of nonbusiness owners who started a business each month.
Design basics	Uses a household-based (from the Census Bureau) sampling frame and rotating sample design: respondents are in the survey for 4 months, out for 8 months, and back in for an additional 4 months.
Frequency	Monthly, longitudinal panel capability upon matching, 1962 to present (matching is imperfect—annual match rates around 70%).
Unit level	Individual, family, and household
Coverage	Civilian noninstitutionalized population ages 16 and over. Survey size is approximately 60,000 households.
Content	Employment (by occupation and industry), indicators for self-employed, unemployment, business ownership, some characteristics of small business employees, earnings, hours of work, age, sex, race, marital status, and educational attainment. Supplemental questions on school enrollment, income, previous work experience, health, employment benefits, and work schedules.
Limitations or lag time	Record matching over time is imperfect.
Accessibility of data	Microdata are publicly available. Data can be accessed electronically.

U.S. CENSUS BUREAU

Census Bureau's Business Register (BR)

Purpose/uses	Provides a comprehensive database of U.S. business establishments and companies for statistical program use. Serves as the master enumeration list for sampling frames drawn for the Census Bureau's firm and establishment surveys, most notably the quinquennial economic census. Source for annual reports providing summary statistics (e.g., number of establishments, payroll, employment) by county and 6-digit NAICS industry.
Design basics	Sample frame draws from the IRS Business Master File, tax return data from Schedule C 1040, SSA data, the economic census, COS, and other Census Bureau business surveys.
Frequency	Establishment listings are initiated and updated continuously with information from Census Bureau and other federal statistical and administrative records programs. Individual data items are updated anywhere from every quarter to every 5 years (with the economic census).
Unit level	Establishment (organized by EIN, enterprise, and alternate reporting units)
Coverage	Employer and nonemployer businesses: 180,000 multiunit companies representing 1.5 million affiliated establishments, 5 million single-establishment companies, and approximately 16.5 million nonemployer businesses.
Content	Business location (mailing and physical address), organization type, EIN, NAICS, LFO code, tax status, employment and payrolls, IRS reported sales and receipts or revenue, assets, interest income, gross rents, parent EIN, activity status, and filing requirement codes.
Limitations or lag time	Geo and industry codes are updated only every 5 years. Lack of detail on small business owners. Accuracy of single versus multiunit identification—and of small multiunit establishment births and deaths—declines between economic censuses; no data on government or farms; ownership links for multistate firms not comprehensive.
Accessibility of data	Information is confidential under Title 13 and Title 26. No public-use data set. Researchers can apply for access to the confidential microdata.

Company Organization Survey (COS)

Purpose/uses	Used to obtain current organization and operating information on multiestablishment firms to maintain and update the BR. Source for CBP Reports.
Design basics	Some multiestablishment companies receive annual mail-out/mail-back surveys. Smaller companies are selected when administrative data indicate a probable organizational change using a probability sampling procedure. About 40,000 multiunit companies with more than 250 employees and about 10,000 smaller multiunit companies are selected on a rotating basis. Content and coverage vary during the 5-year economic census program cycle.
Frequency	Conducted annually since 1974.
Unit level	Multiestablishment firms
Coverage	Cross-sectional survey of multiestablishment companies with payroll (and their establishments), excluding agricultural production companies.
Content	Companies identify establishments (including mailing and physical address) that have been sold, closed, continued, started, and acquired. Businesses are asked about first quarter and annual payroll, employment during the pay period including March 12 for each establishment, large foreign equity positions, and controlling interests held by other domestic or foreign-owned organizations.
Limitations or lag time	Limited scope and rotation of sampled firms over time affect the timeliness and coverage of smaller multiestablishment companies in the BR.
Accessibility of data	No public-use version available. Researcher can apply for access to the confidential microdata, which are typically available for a given year with about an 8-month lag.

U.S. CENSUS BUREAU, continued

The Economic Census

Purpose/uses	Provides a detailed portrait of the economy once every 5 years at both national and local levels. Used to update the Census BR and to produce industry and geographic area series and supplemental surveys of minority- and women-owned businesses.
Design basics	More than 5 million companies are mailed a census form. Large- and medium-size firms, plus all firms known to operate more than one establishment are sent forms. For most very small firms, data from existing administrative records of other federal agencies are used. Data can be linked longitudinally. Geographic detail available varies by sector and can range from state to zip code levels.
Frequency	Data are collected every 5 years (years ending in 2 and 7).
Unit level	Establishment, firm
Coverage	All domestic nonfarm business establishments (not operated by government).
Content	Statistics tabulated for all industries covered include number of establishments, number of employees, payroll, and measure of output (sales, receipts, revenue, value of shipments, or value of construction work done). Additional items are available for certain sectors.
Limitations or lag time	Collected every 5 years, making birth and death coverage incomplete. Nonemployer coverage by sample only (though this can be supplemented using annual nonemployer statistics); no detailed owner characteristics; and no government or farm coverage.
Accessibility of data	Though many publications based on the economic census are readily available, no public-use version of underlying microdata is available. Researcher can apply for access to the confidential microdata.

Longitudinal Research Database (LRD)

Purpose/uses	Provides company-level data that have supported research on employment dynamics as well as on the issues related to productivity, profitability, and the uses of research and development.
Design basics	Data collected primarily using a mail-out/mail-back process. Periodically, visits to key companies are conducted to record the changing nature of R&D activities and any reporting difficulties companies may have, and to determine collectibility of proposed new items. A probability-proportionate-to-size approach is used for selecting individual companies for participation.
Frequency	Underlying survey data collected annually since 1957; however, only data from 1972 forward are included in the database. The R&D database is generally updated annually within 2 years after the survey reference year.
Unit level	Plant
Coverage	Sample size has been approximately 25,000 companies since 1992. In any given year, the number of sampled companies that conduct or sponsor R&D activities is in the 3,500 to 4,000 range. Due to the concentration of R&D activities among the larger companies, most companies with significant R&D activities remain in the sample for a number of years.
Content	*Mandatory items*: federal and company financed R&D, domestic net sales, domestic employment, and R&D by state. *Voluntary items*: information about scientists and engineers employed; basic and applied R&D using federal and company funds; contracted-out, foreign, and budgeted research; R&D by major type of expense and technology area; and energy R&D.
Limitations or lag time	Data historically limited to manufacturing sectors; coverage of firms with fewer than 250 employees is limited; plant-level data are not linked to enterprises.
Accessibility of data	Research is conducted by permanent and specially sworn Census Bureau employees. All current research is done at the Center for Economic Studies.

U.S. CENSUS BUREAU, continued

Longitudinal Busness Database (LBD)

Purpose/uses	Used for researching establishment and firm dynamics (entry and exit) and job flows. Extends the LRD beyond manufacturing sectors
Design basics	Contains annual observations on employment and payroll for all businesses in the U.S. private sector. The database sources are periodic business surveys conducted by the Census Bureau and federal government administrative records. It uses SSEL data and EIN-based year-to-year linking.
Frequency	Annual
Unit level	Establishment and firm
Coverage	Longitudinal data set of all employer business establishments from 1975-2003.
Content	Establishment identifiers, age and tenure, payroll, employment, firm affiliation, name, and location (mailing and physical address) information. Work is ongoing to add payroll employment, location, industry activity, and firm affiliation.
Limitations or lag time	Linkages can be difficult due to inconsistent data formats, changing business ID numbers, and the sheer number of records.
Accessibility of data	None outside Census Bureau. Plans are in place to provide access to microdata available at RDCs after further documentation and quality assurance.

Integrated Longitudinal Business Database (ILBD)

Purpose/uses	Extends LDB coverage to include nonemployer businesses providing research data on firm and job dynamics. The database allows a business's characteristics to be tracked as it transitions from nonemployer to employer status (or vice versa). The ILBD is currently used by researchers working on a wide variety of projects at the Census Bureau's Center for Economic Studies and RDCs.
Design basics	Integrates federal administrative records and survey data in a longitudinal structure. Records are linked by EIN or social security number.
Frequency	Data compiled annually, 1992, 1994 to 2000.
Unit level	Firm
Coverage	All private, nonagricultural, employer, and nonemployer businesses in the United States, currently covering the years 1992 and 1994 to 2000. This amounts to approximately 21 million businesses (over 15 million nonemployers and over 5 million employers).
Content	A range of business characteristics, including location (mailing and physical address) as they transition from nonemployer to employer and visa versa.
Limitations or lag time	Linkages can be difficult due to inconsistent data formats, changing business ID numbers, and the sheer number of records.
Accessibility of data	Data access governed by Title 13 of the U.S. Code. Microdata will be available at RDCs in the near future.

U.S. CENSUS BUREAU, continued

Longitudinal Employer-Household Dynamics (LEHD)

Purpose/uses	A microdata source designed to provide a detailed and comprehensive picture of workers, employers, and their interaction in the U.S. economy. Employment-household linked microdata create opportunities to conduct longitudinal research using on business start-ups early life-cycle dynamics and on local labor market conditions. Used by QWI to measure job churning and in designing the LED program.
Design basics	A set of infrastructure files using administrative data provided by state agencies, enhanced with information from demographic and business surveys and censuses. Uses LBD (linked to household data); federal and state administrative data; core Census Bureau censuses and surveys.
Frequency	Yearly, panel (1992 to 2001).
Unit level	Establishment and household
Coverage	Establishments from about 20 states (about 5 million) and 80 million individual records.
Content	Integrates information about employers (including employment and payroll levels, industry, location, and employment history of employees); employees (including gender, race, foreign-born status, and date of birth); the skill mix of businesses; and employer- level accessions, separations, job creation, and destruction.
Limitations or lag time	Data only reveal workers' quarterly earnings, not work hours. For most workers data are not available on education or family characteristics.
Accessibility of data	Available to authorized users at Census Bureau controlled facilities. No public-use version available. Researcher can apply for access to the confidential microdata.

Survey of Business Owners (SBO)

Purpose/uses	Provides statistics describing the composition of U.S. businesses by gender, race, and ethnicity of owner and sources of financing. Economic policy makers in federal, state, and local governments use SBO data as a source of information on business success and failure rates.
Design basics	Sample frame is based on IRS administrative data. The sample size is typically around 2.3 million businesses.
Frequency	Tied to economic census (every 5 years).
Unit level	Establishment and owner
Coverage	Firms operating during reference year with receipts of $1,000 or more that filed tax forms as individual proprietorships, partnerships, or any type of corporation. Excludes those classified as agricultural production, domestically scheduled airlines, railroads, U.S. Postal Service, mutual funds (except real estate investment trusts), religious grant operations, private households and religious organizations, public administration, and government.
Content	Legal form of organization, receipts, business owner's race (self-identified and allowing for identification of more than one racial group), gender, ethnicity, age, education level, veteran status, and primary function in the business. Also includes family- and home-based businesses, types of customers and workers, and sources of financing for expansion, capital improvements, or start-up.
Limitations or lag time	Infrequent; not longitudinal
Accessibility of data	No public-use version available. Researchers can apply for access to the confidential microdata.

DUN & BRADSTREET (D&B)

Duns Market Identifiers (DMI)

Purpose/uses	Provides basic company data for U.S. business establishments. Used as a sampling frame for a wide variety of government (e.g., the SSBF) and private-sector applications.
Design basics	Data are collected from a wide range of public and private sources including in-person and telephone interviews, government publications, business trade programs, mailings, and applications for credit.
Frequency	Data are updated continuously, albeit in an ad hoc fashion.
Unit level	Establishment and firm
Coverage	U.S. business establishment locations of all sizes and types, including public and private companies, government agencies and contractors, and schools and universities. Includes over 17 million establishments; over 2.9 million private and public companies. Limited to companies with 5 or more employees or sales of $1 million.
Content	Information on owners, sales, employment and legal status, full address, names of executives and titles, corporate linkages, Duns numbers, organization status, marketing information, primary SIC code, and sometimes a NAICS code.
Limitations or lag time	Relies on disparate sources for detecting appearance of new businesses—there are no standard guidelines. There is no distinction between firm and establishments. Information on ownership and small firm characteristics is limited.
Accessibility of data	Microdata available for a fee.

FEDERAL RESERVE BOARD (FRB)

Survey of Small Business Finances (SSBF) [oonducted by NORC]

Purpose/uses	The most comprehensive source of information available on the characteristics of small businesses and their owners, focusing on financial data. Data have been used to prepare the Report to Congress on the Availability of Credit to Small Business every 5 years. Facilitates research on factors affecting prices and availability of credit; characteristics of small businesses and their influence on credit needs; experiences with credit applications; impact of government regulations on credit access; financial and nonfinancial sources used for financing needs. The FRB intends to discontinue the SSBF.
Design basics	About 24,000 firms from D&B are screened for a final sample of 4,240 (for 2003).
Frequency	About every 5 years (1987, 1993, 1998, and 2003); cross-sectional.
Unit level	Firm
Coverage	Nationally representative sample from D&B of firms with fewer than 500 employees; oversamples African American, Asian American, and Hispanic American owned firms.
Content	Income and expenses, assets and liabilities, and financing sources.
Limitations or lag time	Infrequent—last conducted in 2003 with a low response rate—around 33%.
Accessibility of data	Only a small number of authorized staff at NORC and the Federal Reserve System has access to the raw microdata. A public-use version, altered to maintain respondent confidentiality, is available.

GLOBAL ENTREPRENEURSHIP MONITOR CONSORTIUM

Global Entrepreneurship Monitor (GEM)

Purpose/uses	Measure differences in the level of entrepreneurial activity between countries and the relationship between entrepreneurship and national economic growth; uncovers factors that lead to higher levels of entrepreneurship and suggest policies that may enhance levels of entrepreneurial activity. Data have been used to produce an Indicator of Total Entrepreneurial Activity and reports on women and entrepreneurship.
Design basics	Data are collected through a series of coordinated household surveys in an increasing number of countries using a common interview schedule and consolidating and standardizing responses. Adult population surveys range from 1,000 to nearly 27,000 individuals per country—the average sample size is about 2,000.
Frequency	Samples are drawn annually.
Unit level	Household
Coverage	Cross-national assessment of entrepreneurship in 35 countries covering three types of data: adult population surveys, national expert interviews, and standardized cross-national data.
Content	Level of entrepreneurial activity, variance between countries, and change over time; relationship between entrepreneurship and economic growth; how national experts assess entrepreneurial climate in their countries; who becomes an entrepreneur, why and what types of businesses they are creating; and the importance of venture capital and informal finance.
Limitations or lag time	Individual-level data available only after a several year lag (national summary reports are available with a lag of less than a year).
Accessibility of data	GEM consortium members have access to individual-level survey data, interview schedules, data collection procedures, and other material needed for systematic analysis. Public users can view all reports.

INTERNAL REVENUE SERVICE (IRS)

Statistics of Income (SOI)

Purpose/uses	Provides the only publicly available financial information on all corporations. Data products for S-corporations—those with 75 or fewer shareholders—are also available. The SOI provides data annually to BEA on partnerships, as well as producing annual information on nonfarm sole proprietorships from Schedule C data.
Design basics	The survey is based on a stratified probability sample of 130,000 preaudited income tax returns or other forms filed with the IRS.
Frequency	Yearly, cross-section (1990 to 2002).
Unit level	Firm
Coverage	Corporations, S-corporations, partnerships, and nonfarm sole proprietorships.
Content	Net income statements, balance sheets, and tax information by industry, accounting periods, sizes of assets, receipts, and income taxes after credits.
Limitations or lag time	Potential reporting errors and inconsistencies, processing errors, and the effects of any early cutoff of sampling.
Accessibility of data	Though statistics are publicly available, researchers must apply for the access to the confidential microdata.

KAUFFMAN FOUNDATION

Kauffman Firm Survey (KFS)
[with Mathematica Policy Research, Inc.]

Purpose/uses	Produces data on the financial development of new businesses and to track them in the first 4 years of existence. The data set is intended to create a public-use data source that informs policy decisions and academic analysis.
Design basics	Sampled from D&B, a longitudinal survey of the principals of 5,000 firms that started operations in 2004. The survey is oriented primarily to generate data on the financial development of new businesses in their first four years of existence. High-technology businesses are oversampled. Surveys conducted by either telephone or on the Internet
Frequency	Begun in 2005, an annual survey with 3 follow-up panels over the period 2006 to 2008.
Unit level	Owner and firm
Coverage	New businesses starting in year prior to reference year in the United States.
Content	Business characteristics, strategy and innovation, employment, business organization and benefits, business finances, and work behaviors and demographics of owner(s).
Limitations or lag time	D&B sampling frame is limited in ability to quickly incorporate new firms (see D&B).
Accessibility of data	Data not now available; ultimately, publicly available longitudinal data on new firms will be available.

Panel Study of Entrepreneurial Dynamics (PSED)
[with the University of Michigan]

Purpose/uses	A nationally representative database designed to enhance understanding of the business start-up phenomenon. Resulting data are intended to promote research into the business gestation process (i.e., the period before the business actually produces output).
Design basics	A longitudinal sample of U.S. households was contacted to find individuals who were actively engaged in starting new businesses. Those identified as nascent entrepreneurs were included in the group and asked to participate in two follow up interviews (each 12 months apart). Four waves of PSED exist for 1998 to 2003; a new cohort has been developed for interview in the 3 years beginning in 2006.
Frequency	Annual
Unit level	Individual entrepreneurs located by household.
Coverage	Approximately 670 nascent entrepreneurs, identified through a survey of 64,000 U.S. households.
Content	Proportion and characteristics (gender, ethnicity, age, education, household income, and urban context) of the adult population attempting to start new businesses, kinds of activities nascent entrepreneurs undertake during start-up, and proportion and characteristics of start-up efforts that become infant firms.
Limitations or lag time	Does not interview respondents who do not qualify as nascent entrepreneurs when initially selected (doing so could eliminating the need for comparison groups).
Accessibility of data	Data from the PSED are maintained and made available for download by the University of Michigan's Institute for Social Research. Four panels of data are currently available covering the time period 1998 to 2003.

KAUFFMAN FOUNDATION, continued

Kauffman Financial and Business Database (KFBD)

Purpose/uses	To collect financial information on U.S. businesses.
Design basics	Data primarily used for credit scoring purposes are purchased from D&B. Data include recent, detailed financial information. On average the longitudinal database contains complete, consecutive financial statements for a period of 3 years in length.
Frequency	Annual—data are purchased from D&B on a semi-annual basis.
Unit level	Firm
Coverage	The longitudinal file includes data for every year since 1996 and contains more than 2 million records with financial information on more than 500,000 unique firms.
Content	Financial statements for 3-year periods for about 50,000 companies, including annual balance sheet, annual income statement, 14 standard financial ratios, and various firm-level demographic items. Data may be sorted by industry, year started, number of employees, annual sales, minority ownership, and detailed location information.
Limitations or lag time	D&B is limited in its coverage of the newest start-up firms and of self-employed individuals.
Accessibility of data	Data are available for legitimate research from the Kauffman Foundation.

SMALL BUSINESS ADMINISTRATION (SBA)

Statistics of U.S. Businesses (SUSB) (1989 to present)
[compiled by the Census Bureau]

Purpose/uses	An annual series that provides national and subnational data on the distribution of economic activity by size and industry. Provides data on firms, establishments, employment, annual payroll, and estimated receipts (or sales) from which various tables are produced.
Design basics	Data items extracted from SSEL. The annual COS provides individual establishment data for multiestablishment companies. Data for single-establishment companies are obtained from the Annual Survey of Manufactures and Current Business Surveys, as well as from administrative records from IRS, the SSA, and BLS.
Frequency	Annual. Historical comparability is affected by definitional changes in establishments, activity status, and industrial classifications over the period 1988 to 2002.
Unit level	Establishment and firm
Coverage	The 1999 Statistics of U.S. Businesses covers all NAICS industries except crop and animal production, rail transportation, U.S. Postal Service, pension, health, welfare, vacation funds, trusts, estates, agency accounts, private households, and public administration.
Content	Tabulations can be made to estimate employment, annual payroll, number of firms, number of establishments by location and industry categories.
Limitations or lag time	The series excludes data on self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees. There is a 2-year time lag in reporting.
Accessibility of data	Tabulations of data by enterprise size for the country, states, and/or metropolitan statistical area can be accessed for recent years.

Business Information Tracking Series (BITS)
(also known as Longitudinal Establishment and Enterprise Microdata (LEEM))
[constructed by the Census Bureau]

Purpose/uses	To identify firm births and deaths, expansions and contractions, and mergers and acquisitions and for examining job flows.
Design basics	BITS is constructed by longitudinally linking archived SUSB data. The data set currently includes about 13 million establishments.
Frequency	Yearly, panel (1989 to present)
Unit level	Establishment and firm
Coverage	Private-sector establishments (single physical locations) with positive payroll. Same industry coverage as CBP.
Content	Establishment- and firm-level data on annual payroll, 4-digit SIC, location, start year, legal entity, total employment, firm affiliation, census geography, starting year, census file number, and constant firm identifiers (meaning there is no change in the ID even if legal or ownership status changes).
Limitations or lag time	No self-employed; long lag in production (about 2 years); only tracks establishments (not firms), and has no farm coverage.
Accessibility of data	Not publicly available. Must become a sworn Census researcher and use data at Census RDCs.

STANDARD & POOR'S (S&P)

COMPUSTAT

Purpose/uses	Tracks firm level activity for publicly traded, listed firms since 1950. Standardizes financial and accounting statement information on companies around the world for investors. Data used by hedge funds, money managers, analysts, researchers, corporations, and government (the IRS) and regulatory agencies.
Design basics	Database produced by S&P. Reporting units are identified by firm and by 4-digit SIC code and are business or industry segments, defined as a component of an enterprise engaged in providing a product, service, or group of related products or services primarily to customers outside the enterprise for profit.
Frequency	Quarterly (longitudinal since 1980).
Unit level	Firm and industry segment
Coverage	All publicly traded firms in U.S. stock markets (about 65,000 firms).
Content	Data include quarterly and annual income statements, balance sheets, and cash flow statements. Source information includes annual and quarterly SEC filings, 8-K, 20-F and Proxy filings, EDGAR filings and media releases and original annual reports.
Limitations or lag time	By design, limited to publicly traded firms (generally means mature entities).
Accessibility of data	Data available for a fee.

Appendix B

Biographical Sketches of Panel Members and Staff

JOHN HALTIWANGER *(Cochair)* is professor of economics at the University of Maryland. Previously, he served as chief economist at the Bureau of the Census. He currently serves as a research associate at the Center for Economic Studies and at the National Bureau of Economic Research. His recent research has exploited the newly created longitudinal establishment and employer-employee matched databases that have been developed at the Census Bureau. This research centers on the process of job and worker reallocation, retooling and restructuring in the U.S. economy, and the connection of these factors to the business cycle and productivity growth. He is the author of the books *Job Creation and Destruction* and *Labor Statistics Measurement Issues*. He is a member of the American Economic Association, the Econometric Society, and the American Statistical Association. He serves on the editorial board of *Small Business Economics*, the *Journal of Evolutionary Economics*, and the *Journal of Macroeconomics*. At the National Research Council, he is a member of the Committee on National Statistics. He has an Sc.B. in applied mathematics-economics from Brown University and a Ph.D. in economics from the Johns Hopkins University.

LISA M. LYNCH *(Cochair)* is the William L. Clayton professor of international economic affairs and former academic dean of the Fletcher School at Tufts University. She has published over 50 articles and books in her principal fields of research—labor economics, international human resource management, and applied econometrics. She is currently the deputy chair of

the Board of Directors of the Federal Reserve Bank of Boston, chair of the American Economics Association Committee on the Status of Women in the Economics Profession, and executive board member of the Labor and Employment Relations Association. She is the former chief economist of the U.S. Department of Labor, chair of the Federal Economics Statistical Advisory Committee, and coeditor of the *Journal of Labor Economics*. At the National Research Council, she served on the Committee Toward Improved International Labor Standards: Data, Monitoring, and Compliance. She has a B.A. in economics and political science from Wellesley College and M.Sc. and Ph.D. degrees, both in economics, from the London School of Economics.

JOHN M. ABOWD is the Edmund Ezra Day professor of industrial and labor relations at Cornell University, director of the Cornell Institute for Social and Economic Research, distinguished senior research fellow at the Census Bureau, research associate at the National Bureau of Economic Research, and research affiliate at the Centre de Recherche en Economie et Statistique in Paris. His current research focuses on the creation and use of linked, longitudinal data on employees and employers. In his work at the Census Bureau he provides the scientific leadership for the Longitudinal Employer-Household Dynamics Program, which is creating research data integrating demographic surveys, economic surveys, and administrative data. He has an M.A. and Ph.D., both in economics, from the University of Chicago.

PATRICIA M. ANDERSON is professor of economics at Dartmouth College and a faculty research fellow at the National Bureau of Economic Research. Her research includes the use of business microdata, the effects of unemployment insurance payroll taxes on wages and unemployment, the effects on trends in labor force participation and retirement on pensions and social security, and the extent and consequences of job turnover. She is coeditor of the *Journal of Human Resources* and serves on the editorial board of *The B.E. Journals in Economic Analysis and Policy*. She also served as a member of the Federal Economic Statistics Advisory Committee. She has an M.A. and a Ph.D., both in economics, from Princeton University.

MATTHEW BARNES is senior economist in the Better Regulation Executive in the Cabinet Office of the United Kingdom. His current work focuses on the measurement and reduction of regulatory burdens on business and the public sector. This includes work to improve the use of analysis used in policy making and research into new policy options to improve outcomes while managing burdens. He was previously at the UK Office for National

Statistics, where his work focused on productivity, issues concerning the consistency of data from national accounts and labor market statistics sources, and international comparisons of productivity. He has a B.Sc. in economics and finance from the University of York and an M.Sc. in economics from the University of Essex.

STEVEN DAVIS is the William H. Abbott professor of international business and economics at the University of Chicago Graduate School of Business, a research associate with the National Bureau of Economic Research, and a visiting scholar at the American Enterprise Institute. He previously taught at Brown University and the Massachusetts Institute of Technology and served as a consultant and researcher at the Federal Reserve Bank of Chicago. His research focuses on job creation and destruction, tax effects on work activity, the distribution of consumption and earnings, the demand for risky assets, and software pricing and design. He is a member of the American Economic Association, the Econometric Society, and the Society of Labor Economists. He is the coauthor of *Job Creation and Destruction*. He has a B.A. in economics from Portland State University and M.A. and Ph.D. degrees, both in economics, from Brown University.

TIMOTHY DUNNE is senior economic advisor in the Research Department of the Federal Reserve Bank of Cleveland. He was previously the Chong K. Liew professor of economics at the University of Oklahoma. He also earlier served as the director of research in the Office of the Chief Economist at the Census Bureau. His fields of interest focus on applied microeconomics, industrial organization, labor economics, and the economics of technological change. He has a B.A. in economics and history from the College of William and Mary and a Ph.D. in economics from the Pennsylvania State University.

ROBERT M. GROVES is professor of sociology and the director of the Institute for Social Research at the University of Michigan. He has published widely on such topics as the impact of mode of data collection, interviewer behavior, and the use of computer assistance on the quality of survey data. He is the author of *Survey Errors and Survey Costs* and the coauthor of *Nonresponse in Household Surveys*. At the National Research Council, he has served on seven committees and is a former member of the Committee on National Statistics. He has an M.A. in statistics and an M.A. in sociology from the University of Michigan and a Ph.D. in sociology from the University of Michigan.

SUSAN HANSON is a professor in the Graduate School of Geography at Clark University. Her current research focuses on understanding how

gender, geographic opportunity structures, and geographic rootedness affect entrepreneurship in cities, as well as on understanding the emergence of sustainable versus unsustainable practices in urban areas. She is a past president of the Association of American Geographers and a member of the National Academy of Sciences. She has a B.A. in geography from Middlebury College and a Ph.D. in geography from Northwestern University.

CHRISTOPHER MACKIE is a study director with the Committee on National Statistics (CNSTAT) specializing in economic measurement and statistics. In addition to this study, he led a number of other projects, including those that produced the reports, *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*, and *Beyond the Market: Designing Nonmarket Accounts for the United States*. He has also led a number of CNSTAT initiatives related to national health accounting and to data access, sharing, and confidentiality issues. He is the author of *Canonizing Economic Theory*. He has a Ph.D. in economics from the University of North Carolina, and has held teaching positions at the University of North Carolina, North Carolina State University, and Tulane University.

ROBERT H. MCGUCKIN, III, was director of Economic Research at The Conference Board. Previously, he served as Chief of the Center for Economic Studies, U.S. Bureau of the Census. His most recent work focused on international comparisons of productivity and research and development, diffusion of information and communication technology, mergers and acquisitions, and economic and business cycle indicators. He received his B.A. in Mathematics from Ithaca College in 1965, his M.A. in Economics from the State University of New York at Buffalo in 1968, and his Ph.D. in Economics from the State University of New York at Buffalo in 1970. He was a member of the American Economic Association, the National Association of Business Economists, and the American Statistical Association.

PAUL D. REYNOLDS is a distinguished visiting professor of management in the College of Business Administration of Florida International University. He also serves as director of the Entrepreneurial Research Institute of the Eugenio Pino and Family Global Entrepreneurship Center. He has been the coordinating principal investigator on a number of research programs, including the Panel Study of Entrepreneurial Dynamics and the Global Entrepreneurship Monitor program; both programs have involved consortia of three dozen units and over 100 scholars. Since 1995 he has served as professor of entrepreneurship at the London Business School and as professor of entrepreneurial studies at Babson College; in prior years he taught at a number of universities, including 20 years at the University of Minnesota.

He is the author of *The Entrepreneurial Process* (with Sammis White), *Ethics and Social Science Research, Ethical Dilemmas and Social Science Research*, and *A Primer in Theory Construction*. He has an M.B.A. in business, an M.A. in psychology, and a Ph.D. in sociology, all from Stanford University.

MARK J. ROBERTS is professor of economics at the Pennsylvania State University. His research focuses on applied microeconomics and industrial organization. He serves on the American Economic Association Census Advisory Committee, the Executive Committee of the Conference on Research in Income and Wealth, and is a research associate for the National Bureau of Economic Research. He is the coeditor of *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*. He has an A.B. in economics and journalism from Syracuse University and an M.A. and a Ph.D., both in economics, from the University of Wisconsin.

NIELS WESTERGAARD-NIELSEN is director of the Center for Corporate and professor of economics at the Aarhus School of Business in Aarhus, Denmark. He also serves as a research fellow at the Institute for the Study of Labor in Bonn, Germany, and as a member of the Personnel Economics Group of the National Bureau of Economic Research. He is country coordinator in a project conducted by Russell Sage Foundation on low-wage work. His research focuses on wage formation at the individual and firm level, matched worker and firm data, personnel economics, and health economics. He is the author of a number of journal articles utilizing the Danish registers for analyses of individual behavior in the traditional areas of economics as well as psychiatry. He is coauthor of *The Danish Economy: An International Perspective*. He serves as a member of the Scientific Committee of the Work Environment Program, Ministry of Employment, and is a member of the Danish Government Committee on Work Life Balance. He has an M.Sc. in economics from the University of Aarhus and a Ph.D. from the University of Copenhagen.

KIRK WOLTER is senior fellow and director of the Center for Excellence in Survey Research at the National Opinion Research Council, where he previously served as senior vice president for statistics and methodology. He is also professor of statistics at the University of Chicago. During his career, he has led or participated in the design of many of America's largest information systems, including the Current Business Surveys, the Current Employment Statistics program, the Current Population Survey, the 1980 and 1990 decennial censuses, the National Longitudinal Survey of Youth 1997, and the National Resources Inventory. He is a fellow of the Ameri-

can Statistical Association, an elected member of the International Statistical Institute, and past president of the International Association of Survey Statisticians and of the Survey Research Methods Section of the American Statistical Association. He has an M.A. and a Ph.D. in statistics, both from Iowa State University.

Index

C

D

E

J

K

L

M

N

O

P

T

U