

IZA DP No. 4052

**Entrepreneurship, Self-Employment and Business Data:
An Introduction to Several Large, Nationally-Representative
Datasets**

Robert W. Fairlie
Alicia M. Robb

March 2009

Entrepreneurship, Self-Employment and Business Data: An Introduction to Several Large, Nationally-Representative Datasets

Robert W. Fairlie

*University of California, Santa Cruz
and IZA*

Alicia M. Robb

University of California, Santa Cruz

Discussion Paper No. 4052
March 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Entrepreneurship, Self-Employment and Business Data: An Introduction to Several Large, Nationally-Representative Datasets

Only a few large, nationally-representative datasets include information on both the owner and the business. We briefly describe several of the most respected and up-to-date sources of data on entrepreneurs, the self-employed, and small businesses. More information including estimates of recent trends in business ownership and performance (e.g. survival rates, sales, employment, payroll, profits and industry) from these datasets is contained in Fairlie and Robb (2008).

JEL Classification: L26, Y1

Keywords: entrepreneurship, self-employment, business owners, small businesses, data

Corresponding author:

Robert W. Fairlie
Department of Economics
University of California
Engineering 2 Building
Santa Cruz, CA 95064
USA
E-mail: rfairlie@ucsc.edu

Introduction

A major challenge for researchers in the area of entrepreneurship, self-employment and small business research is finding data. It is difficult to find detailed information on both the characteristics of the owner and the characteristics of the firm. This is important if we want to learn more about the relationship between owner characteristics, such as race, gender, education and prior work experience, and business performance. In particular, few datasets have both large enough samples of underrepresented groups in business, such as minorities or women, and detailed information on businesses characteristics and outcomes. To be sure, there are many large nationally representative datasets based on household surveys such as the Census and Current Population Survey (CPS), but these datasets include only limited information on the owner's business. Household surveys typically include only the owner's income from the business and the industry of the business. Business-level datasets on the other hand include more information on the business, but typically include very limited information, if any, on the characteristics of the owners.

In this paper, we briefly discuss several datasets that include information on the owner and at least some information on the business. We emphasize datasets with information on race, but these datasets also include detailed information on other owner characteristics. Information on the owner is essential for understanding the exogenous determinants of entrepreneurship and business success. In Fairlie and Robb (2008), we use some of these datasets to provide a comprehensive analysis of recent trends in total and racial patterns in business ownership and performance in the United States. We also

present an extensive analysis of the owner and firm characteristics that are associated with business success, as well as the causes of racial differences in business outcomes.

The Survey of Business Owners and the Survey of Minority-Owned Business Enterprises

The Survey of Business Owners (SBO) is conducted by the U.S. Census Bureau every five years to collect statistics that describe the composition of U.S. businesses by gender, race, and ethnicity. This survey was previously conducted as the Survey of Minority- and Women-Owned Business Enterprises (SMOBE/SWOBE). Published estimates from the SMOBE and SBO are the most commonly cited and used estimates of the state of minority-owned businesses. The data are compiled from several sources: IRS business tax returns, other Economic Census reports. (e.g., Annual Survey of Manufacturers; Annual Retail Trade Survey), Social Security information on race and Hispanic or Latino origin; and a mailout/mailback survey. The universe for the most recent survey is all firms operating during 2002 with receipts of \$1,000 or more that filed tax forms as individual proprietorships, partnerships, or any type of corporation.¹ Businesses that are 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 are excluded.

The SMOBE and SBO data have undergone several major changes over time including the addition of C corporations and the removal of firms with annual receipts between \$500 and \$1,000 starting in 1997. Table 1 includes a summary of the major

¹ Sole proprietorships complete a 1040C form, partnerships complete a 1065 form, S corporations

changes in the SMOBE/SBO data from 1982 to 2002. Additional changes were made over time, which are described in more detail in Fairlie and Robb (2008). Also, see Fairlie and Robb (2008) for a new compilation of estimates of total and racial business performance from the SMOBE and SBO that makes these data as comparable as possible over the time period 1982-2002.

These surveys provide the most comprehensive data available on businesses by the race, ethnicity, and gender of the owners. Business ownership is defined as having 51 percent or more of the stock or equity in the business. Business ownership was categorized by: Gender (Male; Female; or Equally Male-/Female-Owned); Ethnicity (Hispanic, non-Hispanic); and Race (White; Black or African American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander). For the first time, respondents could choose multiple race groups in the 2002 SBO survey.

Aggregate publications from this survey are available for each racial group, Hispanics, and women. Separate reports are not available for white firms, but estimates for non-Hispanic white firms and all white firms are reported in 1997 and 2002, respectively. Although not directly comparable over time, the number of non-Hispanic white or all white firms can be estimated by subtracting out all minority firms or Hispanic firms in other years.

The confidential microdata underlying these aggregate publications are available to approved researchers at the Center for Economic Studies.² Alternatively, special cross-tabulations of these data can be purchased from the U.S. Census Bureau. Prior to 2002, The SMOBE data contained information on businesses including legal form of

complete a 1120S form, and C corporations complete a 1120 form.

² See www.ces.census.gov for information on how to apply for access to these data.

organization (i.e. sole proprietorships, partnerships, S corporations, and, beginning in 1997, C Corporations), industry, location, employment size, receipts size, and payroll. The 2002 SBO contains all of this information except legal form of organization.

The Characteristics of Business Owners

The Characteristics of Business Owners (CBO) is one of the only datasets that includes detailed owner and firm characteristics for a large sample of business owners. The 1992 Characteristics of Business Owners (CBO) Survey is the third survey of its kind conducted by the U.S. Census Bureau. The first two surveys were conducted for 1982 and 1987. The 1997 CBO was cancelled, and the 2002 CBO (which is the SBO described above) contains substantially less information on the owner and business and has only recently accepted applications for restricted-access use by the U.S. Census Bureau. Minorities and women are oversampled in the 1992 CBO to allow researchers to more reliably study these businesses and business owners. The sample for the CBO is a sub-sample of the SMOBE and drawn from the businesses that responded to the SMOBE (the 1992 SMOBE had a 77 percent response rate).

The 1992 CBO survey was conducted by the U.S. Census Bureau to provide economic, demographic and sociological data on business owners and their business activities (see U.S. Census Bureau 1997, Bates 1990, Headd 1999, and Robb 2000 for more details on the CBO). The survey was sent to more than 75,000 firms and 115,000 owners who filed an IRS form 1040 Schedule C (individual proprietorship or self-employed person), 1065 (partnership), or 1120S (subchapter S corporation).³ Only firms

³ C corporations were not included. C corporations as a tax filing status, however, are becoming less popular relative to S corporations due to changes in tax laws (Headd 1999).

with \$500 or more in sales were included. The businesses included in the CBO represent nearly 90 percent of all businesses in the United States (U.S. Census Bureau 1997).

Response rates for the firm and owners surveys were approximately 60 percent. Sample weights that adjust for survey non-response are available (Headd 1999).

The CBO is unique in that it contains detailed information on both the characteristics of business owners and the characteristics of their businesses. For example, owner characteristics include education, detailed work experience, hours worked in the business and how the business was acquired. Business characteristics include closure, profits, sales, employment and industry. Most business characteristics refer to 1992, with the main exception being closure, which is measured over the period 1992 to 1996. Additional advantages of the CBO over other nationally representative datasets for this analysis are the availability of measures of business ownership among family members and the large oversample of minority-owned businesses. In particular, the CBO contains rare or unique information on business inheritances, business ownership among family members, prior work experience in a family member's business, and prior work experience in a business whose goods/services were similar to those provided by the owner's business. In Fairlie and Robb (2008), the CBO is used to conduct a detailed analysis of the determinants of several business outcomes, such as closure rates, sales, profits, and employment size.

Unfortunately, future CBO surveys are likely to contain substantially less information on owner and firm characteristics. For example, the survey used for the 2002 SBO does not contain information on key variables of interest such as the amount of startup capital, family business experience, prior work experience, and profits. The CBO

microdata can also be used to create a more selective sample that includes firms that meet a minimum weeks and hours restriction (Fairlie and Robb 2008). Imposing a weeks and hours restrictions rules out very small-scale business activities such as casual, consulting or side businesses owned by wage and salary workers.

Given the detailed information on both owner and business characteristics, the oversamples of minority-owned businesses, and availability since 1982, it is surprising that the CBO microdata have only been used by a handful of researchers to study minority-owned businesses (see Bates 1997 and many other studies; Christopher, 1993, 1998; Kijakazi, 1997, and Robb 2000, 2002). The lack of use appears to be primarily due to difficulties in accessing and reporting results from these confidential, restricted-access data. All research using the CBO must be conducted in a Census Research Data Center or at the Center for Economic Studies (CES) after approval by the CES and IRS, and all output must pass strict disclosure regulations.

The Current Population Survey

The Current Population Survey (CPS) provides the most up-to-date estimates of self-employed business ownership rates in the United States. Unlike business-level data, the CPS includes information on wage and salary workers allowing one to easily calculate the percent of the working population that owns a business, referred to as the self-employed business ownership rate. The CPS, conducted by the U.S. Census Bureau and the Bureau of Labor Statistics, is representative of the entire U.S. population and interviews more than 50,000 households and 130,000 people per month. It contains detailed information on labor force and demographic characteristics.

New estimates of business ownership rates from CPS microdata are presented in Fairlie and Robb (2008). These estimates improve on published estimates from the same source by the Bureau of Labor Statistics (BLS). Regularly published estimates from the BLS, such as those reported in *Employment and Earnings*, do not include incorporated business owners, which represent roughly one third and a growing share of all business owners.⁴ Using CPS microdata, owners of both unincorporated and incorporated businesses can be identified.

These data also provide a different, individual-based, representation of recent trends in minority business ownership than estimates of the number of businesses from the SBO/SMOBE. Estimates from the CPS do not suffer from two problems facing the SMOBE -- the scope of businesses included in the SMOBE has changed over the past two decades and the data include a large number of side or "casual" businesses owned by wage and salary workers or individuals who are not in the labor force.⁵ The CPS microdata include all individuals who identify themselves as self-employed in their own not incorporated or incorporated business on their main job, and thus capture only primary business owners.

The CPS is also advantageous for calculating business ownership rates because it provides estimates of the size of the population employed in the wage and salary sector. This allows for the use of the same definitions of race, ethnicity and labor force attachment for estimating both the number of self-employed business owners (numerator)

⁴ See Hipple (2004) for recent estimates of the number of unincorporated and incorporated business owners.

⁵ Estimates from the 1992 CBO, which is drawn from the SMOBE, indicate that 44.2 percent of owners in the survey report that their businesses provided less than 25 percent of their total personal income (U.S. Census Bureau 1997). This estimate, however, does not include the hours worked restriction used in our analyses of the CBO microdata.

and the total number of wage and salary workers and self-employed business owners (denominator).

The Current Population Survey (CPS) Outgoing Rotation Group (ORG) files are used to calculate the self-employment series reported in Fairlie and Robb (2008). The ORG files contain annual samples that are roughly three times larger than those from a monthly CPS, such as the commonly used March Annual Demographic Files. The large sample sizes are useful for estimating trends in business ownership rates for smaller demographic groups and provide more precise estimates of the number and rates of self-employed business ownership for all groups.

In the CPS, self-employed business owners are typically defined as those individuals who identify themselves as self-employed in their own not incorporated or incorporated business on the class of worker question. The class of worker question refers to the job with the most hours during the reference week. The sample can also be restricted to include only individuals ages 16 and over who worked at least 15 hours during this week. The hours restriction rules out very small-scale business activities. As in many studies of self-employment, agricultural industries can also be excluded.

Individual- Versus Business-Level Data

Estimates of the level of minority business activity in the United States can be generated from both individual-level datasets such as the CPS and business-level datasets such as the SBO/SMOBE and compared with each other. Of course, estimates of the total number of business owners and the total number of businesses are not perfectly comparable. Multiple businesses owned by one individual count only once in individual-

level data and businesses with multiple owners count only once in business-level data. These discrepancies are relatively minor, however. Estimates from the 1992 CBO indicate that the total number of business owners is only 12 percent larger than the total number of businesses (U.S. Census Bureau 1997). Similarly, Boden and Nucci (1997) find that less than 3 percent of small business records in the CBO pertain to owners of multiple businesses.

There are several additional reasons for discrepancies in the total number of self-employed business owners and the total number of businesses that are due to measurement issues, some of which are specific to the CPS and SBO/SMOBE. The most important is that business ownership in the CPS refers to the person's main job activity. In the 2002 SBO and 1997 SMOBE, all businesses resulting from filing tax forms to the IRS with at least \$1,000 in annual sales are included. Thus, side businesses owned by wage and salary workers are counted in the SBO/SMOBE data, but not in the CPS data. The focus on main job activity in the CPS may only represent a small part of the discrepancy, however. Headd (2005) finds that in the 2004 CPS, roughly 500,000 of the individuals not reporting self-employment as their main job activity report self-employment in their secondary occupation.

As discussed previously, we impose hours worked restrictions to the CPS sample to rule out small scale businesses. Excluding these businesses further limits the comparability of our CPS estimates to the SBO/SMOBE estimates. The SBO/SMOBE data do not impose any restrictions on the size of the business other than the annual sales restriction. Using the CBO microdata, however, we are able to impose restrictions on hours and weeks worked to remove small-scale businesses.

Another cause of the discrepancy is that the SBO/SMOBE data refer to businesses that existed at any point in the calendar year. The CPS instead refers to self-employed business ownership at the time of the survey. Thus, the CPS is likely to capture fewer business owners because there is a considerable amount of volatility in business ownership. The CPS and SBO/SMOBE may also differ in how likely they are to capture some occupations, such as sales and real estate agents. These individuals may report working for an employer instead of self-employment on the CPS questionnaire, even when they file as sole proprietors (Headd 2005). In a recent preliminary study, Bjelland, et al. (2006) find that a large percent of individuals in the CPS are not found in the Census Bureau's comprehensive dataset of businesses, the Business Register and vice versa. Their preliminary matching exercise, however, only includes unincorporated business owners, sole proprietorships, and businesses with positive earnings.

Overall, individual- and business-level data capture different measures of business activity. There are advantages and disadvantages to both types of data for studying racial business patterns. We present estimates from both sources and CBO microdata in Fairlie and Robb (2008). The CBO microdata capture detailed information on owner and business characteristics and allow us to address some of the disadvantages of using business-level data based on tax filings.

Additional Data Sources for Studying Business Owners

There are several additional nationally representative datasets that can be used to study owner characteristics and their businesses. These datasets, however, are limited in size, scope, and/or timeliness compared with the CPS, SBO/SMOBE and CBO. We

briefly discuss a few of the most commonly used datasets for studying minority entrepreneurship which also have detailed owner information.

One of the only business-level datasets that provides information on the owner is the Survey of Small Business Finances (SSBF), which is conducted by the Board of Governors of the Federal Reserve System every five years. The 2003 SSBF contains a large sample of 4,240 for-profit, non-governmental, non-agricultural businesses with fewer than 500 employees.⁶ The SSBF provides detailed information on many owner and firm characteristics, including credit histories, recent borrowing experiences, balance sheet data, and sources of financial products and services used (see Mach and Wolken 2006 and www.federalreserve.gov/ssbf for more information). One limitation to the SSBF for research on underrepresented groups, however, is that it contains relatively small minority business sample sizes.⁷

A new longitudinal business-level dataset from the Kauffman Foundation is the Kauffman Firm Survey (KFS). The KFS contains information on nearly 5,000 firms that started in 2004 and tracks them over time. There are currently four years of data available. The Kauffman Foundation expects to follow this cohort of new firms for a total of eight years. The sample sizes are sufficiently large for analyses of several minority groups (Asians, Hispanics, and blacks). The KFS includes detailed information

⁶ The underlying sample frame for the SSBF was the Dun's Market Identifiers™ (DMI) file produced by the Dun & Bradstreet Corporation. The DMI data may undercount minority businesses (Haggerty, et al. 1999). To address this concern the SSBF relies on its own screener question on minority ownership.

⁷ The 2003 sample includes 119 black-owned firms, 170 Asian-owned firms, and 149 Hispanic-owned firms (Mach and Wolken 2006). The 1998 survey includes larger minority sample sizes with 273 black-owned firms, 214 Asian-owned firms, and 260 Hispanic-owned firms (Bitler, Robb, and Wolken 2001)

on business outcomes in addition to owner characteristics and information on sources of financial capital.

There are also several commonly-used household surveys providing individual-level data on self-employed business owners in addition to the CPS.⁸ The largest source of data on self-employed business owners is the Census of Population 5% Public Use Microdata Sample (PUMS). The 5% PUMS is calculated every 10 years with observations for 5 percent of the U.S. population. Although Census microdata is available as far back as 1850, the 1910 Census was the first census to include information on self-employment (see www.census.gov and www.ipums.org/usa/). The American Community Survey (ACS) replaces the detailed information in the Census and is released every year (see www.census.gov/acs). The information on owner and business characteristics available in the Census and ACS is similar to the CPS.

Estimates of self-employed business owners can also be obtained from panel data allowing for the analysis of entry and exit over time. Although the CPS described above is typically used as a cross-sectional dataset, panel data can be created by linking consecutive months or years of the survey. By linking consecutive surveys, a two-month up to a two-year panel can be created for a very large sample of individuals. With the detailed information on demographics, human capital and employment characteristics available in the CPS, an extensive analysis of the determinants of business creation and exit at the individual level can be performed. Estimates of entrepreneurship or business

⁸ Information on nascent entrepreneurs can be obtained from the Panel Study of Entrepreneurial Dynamics (PSED). See <http://projects.isr.umich.edu/psed/> for more information.

creation and more discussion of the data are reported in "The Kauffman Index of Entrepreneurial Activity" (Fairlie 2008).⁹

The Survey of Income and Program Participation (SIPP), conducted by the U.S. Census Bureau, provides monthly data for a large sample of individuals (see www.sipp.census.gov/sipp). The Panel Study of Income Dynamics (PSID) provides annual data for a large sample of individuals starting in 1968. The dataset, which contains an oversample of minority households, is conducted at the Survey Research Center, Institute for Social Research, University of Michigan (see <http://psidonline.isr.umich.edu/> for more details). The National Longitudinal Surveys (NLS) are a set of surveys containing detailed longitudinal information on demographic characteristics and labor market activities for several age cohorts of the population (see www.bls.gov/nls/). The National Longitudinal Survey of Youth 1979 (NLSY79) is perhaps the most widely used NLS. It contains a nationally representative sample of individuals between the ages of 14 to 22 when they were first interviewed in 1979 with oversamples of black and Latino youth (see Center for Human Resource Research 2005 for more information).¹⁰ In most cases, these longitudinal datasets provide more detailed information on owner characteristics than the CPS, but they have much smaller sample sizes and are generally not as representative of the current U.S. population.

Summary

Only a few large, nationally-representative datasets include information on both the owner and the business. We briefly describe several of the most respected and up-to-

⁹ The microdata can be obtained at www.kauffman.org.

¹⁰ Also, see Fairlie (2005) for review of the literature using the NLSY79 to study self-employment.

date sources of data on entrepreneurs, the self-employed, and small businesses. More information including estimates of recent trends from these datasets is contained in Fairlie and Robb (2008).

References

- Bates, Timothy. 1990. "The Characteristics of Business Owners Data Base," Journal of Human Resources 25(4): 752-56.
- Bates, Timothy. 1997. Race, Self-Employment & Upward Mobility: An Illusive American Dream, Washington, D.C.: Woodrow Wilson Center Press and Baltimore: John Hopkins University Press.
- Bitler, Marianne, Alicia Robb, and John Wolken, 2001. "Financial Services Used by Small Businesses: Evidence from the 1998 Survey of Small Business Finances" Federal Reserve Bulletin, Vol. 87, April 2001.
- Bjelland, Melissa, John Haltiwanger, Kristin Sandusky, and James Spletzer. 2006. "Reconciling Household and Administrative Measures of Self-Employment and Entrepreneurship," U.S. Census Bureau Working Paper.
- Boden, Richard J. and Alfred R. Nucci. 1997. "Counting the Self-Employed Using Household and Business Sample Data," Small Business Economics, 9 (5): 427-436.
- Center for Human Resource Research. 2005. NLSY79 Users' Guide, Columbus, Ohio: The Ohio State University.
- Christopher, Jan. 1993. Empirical Analyses of the Factors Influencing Business Performance. Ph.D. Dissertation, Howard University.
- Fairlie, Robert W. 2005. "Self-Employment, Entrepreneurship and the National Longitudinal Survey of Youth," *Monthly Labor Review, Special Issue Commemorating the 25th Anniversary of the NLSY79*, 128(2): 40-47.
- Fairlie, Robert W. 2008. The Kauffman Index of Entrepreneurial Activity: 1996-2007. Ewing Marion Kauffman Foundation: Kansas City.
- Fairlie, Robert W., and Alicia M. Robb. 2008. *Race and Entrepreneurial Success: Black-, Asian-, and White-Owned Businesses in the United States*. Cambridge: MIT Press.
- Haggerty, Catherine, Karen Grigorian, Rachel Harter, and John Wolken. 1999. "The 1998 Survey of Small Business Finances: Sampling and Level of Effort Associated with Gaining Cooperation from Minority-Owned Businesses." NORC and Federal Reserve Working Paper.
- Headd, Brian. 1999. "The Characteristics of Business Owners Database, 1992." U.S. Census Bureau, Center for Economic Studies, Working Paper Series, CES-WP-99-8.
- Headd, Brian. 2005. "Measuring Microenterprise: Data on Self-Employment and Nonemployers," Working Paper.

Hipple, Steven. 2004. "Self-Employment in the United States: an Update," Monthly Labor Review, 127(7): 13-23.

Kijakazi, K. 1997. *African-American Economics Development and Small Business Ownership*. New York and London: Garland Press.

Mach, Traci L., and John D. Wolken. 2006. " Financial Services Used by Small Businesses: Evidence from the 2003 Survey of Small Business Finances," Federal Reserve Bulletin, Vol. 87, October 2006, 167-195.

Robb, Alicia. 2000. "The Role of Race, Gender, and Discrimination in Business Survival", Doctoral Dissertation, University of Michigan Press.

Robb, Alicia. 2002. "Entrepreneurship: A Path for Economic Advancement for Women and Minorities?," Journal of Developmental Entrepreneurship, Volume 7, No. 4.

U.S. Census Bureau. 1997. 1992 Economic Census: Characteristics of Business Owners. Washington, D.C.: U.S. Government Printing Office.

Table A.1
Major Changes in Survey Methodology
Survey of Minority-Owned Business Enterprises (1982-1997) and Survey of Business Owners (2002)

	1982	1987	1992	1997	2002
Dataset	SMOBE	SMOBE	SMOBE	SMOBE	SBO
Types of businesses included					
Sole proprietorships	Yes	Yes	Yes	Yes	Yes
Partnerships	Yes	Yes	Yes	Yes	Yes
S-Corporations	Yes	Yes	Yes	Yes	Yes
All other corporations	No	No	No	Yes	Yes
Minimum business receipts	None	\$500	\$500	\$1,000	\$1,000

Source: U.S. Census Bureau (2005). For more details of changes see Table A.2.