

APPENDIX H.

Description of Data Sources for Marketplace Analyses

To perform the analyses in Appendices D through G, BBC used data from a range of sources, including U.S. Census Bureau Public Use Micro-samples (PUMS) for the 1980 and 2000 Census and the 2008 American Community Survey (ACS). BBC also analyzed data from the Federal Reserve Board's 1998 and 2003 Survey of Small Business Finances (SSBF) and the 2002 Survey of Business Owners (SBO), conducted by the U.S. Census Bureau.

PUMS Data

Focusing on the construction and engineering industries, the study team used PUMS data to analyze:

- Demographic characteristics;
- Measures of financial resources;
- Educational attainment; and
- Self-employment (business ownership).

PUMS data offer several features ideal for the analyses reported in this study, including historical cross-sectional data, stratified national and state-level samples, and large sample sizes that enable many estimates to be made with a high level of statistical confidence, even for subsets of the population (e.g., ethnic and occupational groups).

BBC obtained selected Census and ACS data via the Minnesota Population Center's Integrated Public Use Micro-data Series (IPUMS). The IPUMS program provides online access to customized, accurate data extracts.¹ For the analyses contained in this report, BBC used the 1980 and 2000 Census 5 percent samples and the 2008 ACS 1 percent sample.

¹ Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Data for 2000. The 2000 U.S. Census 5 percent sample contains 14,081,466 observations. When applying the Census person-level population weights, this sample represents 281,421,906 people in the United States. The 2000 Oklahoma sub-sample contains 173,843 individual observations, weighted to represent 3,450,058 people.

Categorizing individual race/ethnicity. To define race/ethnicity for the 2000 Census dataset, BBC used the IPUMS race/ethnicity variables RACED and HISPAN to categorize individuals into one of seven groups:

- Non-Hispanic white;
- Hispanic American;
- African American;
- Asian-Pacific American;
- Subcontinent Asian American;
- Native American; and
- Other minority (unspecified)

An individual was considered “non-Hispanic white” if they did not report Hispanic ethnicity and indicated being white only — not in combination with any other race group. All self-identified Hispanic individuals (based on the HISPAN variable) were considered Hispanic American, regardless of any other race or ethnicity identification.

For the five other racial groups, an individual’s race/ethnicity was categorized by the first (or only) race group identified in each possible race-type combination. BBC uses a rank ordering methodology similar to the 2000 Census data dictionary rank ordering. An individual who identified multiple races was placed in the reported race category that ranked highest in BBC’s ordering. African American is first, followed by Native American, then Asian-Pacific American and finally Subcontinent Asian American. For example, if an individual identified “Korean,” this person belongs in the Asian-Pacific American category; if the individual identified “Korean” in combination with “Black,” the individual is considered African American. Again, Hispanic origin overshadowed any other race group identification.

- The Asian-Pacific American category included the following race/ethnic groups: Cambodian, Chamorro, Chinese, Filipino, Guamanian, Hmong, Indonesian, Japanese, Korean, Laotian, Malaysian, Native Hawaiian, Samoan, Taiwanese, Thai, Tongan, and Vietnamese. This category also included other Polynesian, Melanesian and Micronesian races as well as individuals identified as Pacific Islanders.
- The Subcontinent Asian American category included these race groups: Asian Indian (Hindu), Bangladeshi, Pakistani, and Sri Lankan. Individual who identified as “Asian,” but were not clearly categorized as Subcontinent Asian were put into the Asian-Pacific group. (Overall, about eight in ten Asians counted in the 2000 Census in Oklahoma were Asian-Pacific Americans.)

- American Indian, Alaska Native and Latin American Indian groups are considered Native American.
- If an individual is identified with any of the above groups and an “other race” group, the individual is categorized into the known category. Individuals identified as “other race” or “white and other race” are categorized as “other minority.”

In some cases where sample sizes were small, BBC combined the Asian Pacific American, Subcontinent Asian American and “other minority” category into a single “other” category.

Business ownership. BBC used the Census detailed “class of worker” variable (CLASSWKD) to determine self-employment. Individuals were classified into eight categories.

- Self-employed for a non-incorporated business;
- Self-employed for an incorporated business;
- Wage or salary employee for a private firm;
- Wage or salary employee for a non-profit organization;
- Employee of the Federal government;
- Employee of a State government;
- Employee of a local government; or
- Unpaid family worker.

BBC counted individuals who reported self-employment — either for an incorporated or a non-incorporated business — as business owners.

Defining selected industry sectors. The construction sector is defined using the 2000 Census code for the industry, 077, which is equivalent to the 1997 NAICS code 23. The architectural, engineering and related services industry is Census code 729; corresponding to 1997 NAICS code 5413.²

² Some variable codes changed during the time of BBC’s analyses. For example, coding for the industry variable (IND) changed slightly (industry code 077 became 770) from when the study team began analyzing the Census data. Such changes, however, did not affect the comparability of variables, and all conventions presented in this appendix still applied.

Education variables. BBC used the variable indicating respondents' highest level of educational attainment (EDUCD) to classify individuals into four categories:³

- Less than high school;
- High school diploma;
- Some college or associate's degree; and
- At least a bachelor's degree.

Definition of workers. The universe for the class of worker, industry and occupation variables includes workers 16 years of age or older who are gainfully employed and those who are unemployed but seeking work. "Gainfully employed" means that the worker has reported an occupation as defined by the Census code, OCC.

1980 Census data. BBC compared 2000 Census data with data for the 1980 Census to analyze changes in worker demographics, educational attainment and business ownership over time. The 1980 Census 5 percent sample includes 11,343,120 observations weighted to represent 226,862,400 people. The sample includes 151,298 observations in Oklahoma, weighted to represent 3,025,960 people.

A number of changes in variables and coding took place between the 1980 and 2000 Censuses.

Changes in race/ethnicity categories between censuses. Figure H-1 lists the seven BBC-defined race/ethnic categories with the corresponding 1980 and 2000 Census race groups. Combinations of race types are available in the 2000 Census but not in the 1980 Census. The Bureau of the Census introduced categories in 2000 representing a combination of race types to allow individuals to select multiple races when responding to the questionnaire.

For example, an individual who is primarily white with Native American ancestry could choose the "white and American Indian/Alaska Native" race group in 2000. However, if the same individual received the 1980 Census questionnaire, she would need to choose a single race group — either "white" or "American Indian/Alaska Native." Such a choice would ultimately depend on unknowable factors including how strongly the individual identifies with her Native American heritage.

In addition, the data analyst does not have information about the proportions of individual ancestry in 2000 and can only know that a particular individual has mixed ancestry. The variability introduced by allowing multiple race selection complicates direct comparisons between years with respect to race/ethnicity. Even so, 98 percent of survey respondents in 2000 indicated a single race.⁴

³ In the 1940-1980 samples, respondents were classified according the highest year of school completed (HIGRADE). In the years after 1980, this method was used only for individuals who did not complete high school, and all high school graduates were categorized based on the highest degree earned (EDUC99). The EDUC variable merges two different schemes for measuring educational attainment by assigning to each degree the typical number of years it takes to earn.

⁴ Grieco, Elizabeth M. & Rachel C. Cassidy. "Overview of Race and Hispanic Origin," *Census 2000 Brief*, March 2001, page 3.

Figure H-1.
BBC race/ethnic categories compared with Census race and Hispanic Origin
survey questions, 1980 and 2000

BBC-defined race/ethnic categories	2000 Census	1980 Census
African American	Hispanic origin: no Race: Black/Negro alone or in combination with any other non-Hispanic group	Hispanic origin: no Race: Black/Negro
Asian-Pacific American	Hispanic origin: no Race: Chinese, Taiwanese, Japanese, Filipino, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, Indonesian, Malaysian, Samoan, Tongan, Polynesian, Guamanian/Chamorro, Native Hawaiian, Pacific Islander, Micronesian, Melanesian, or other Asian, either alone or in combination with any non-Hispanic, non-Black, or non-Native American groups.	Hispanic origin: no Race: Chinese, Japanese, Filipino, Korean, Vietnamese, Native Hawaiian, Pacific Islander or other Asian
Subcontinent Asian American	Hispanic origin: no Race: Asian Indian, Bangladeshi, Pakistani or Sri Lankan, alone or in combination with white or other groups only	Hispanic origin: no Race: Asian Indian
Hispanic American	Hispanic origin: yes Race: any race groups, alone or in combination with other groups	Hispanic origin: yes Race: any or Hispanic origin: no Race: Spanish
Native American	Hispanic origin: no Race: American Indian or Alaskan Native tribe identified alone or in combination with any non-Hispanic, non-Black group.	Hispanic origin: no Race: American Indian/Alaska Native
Other minority group	Hispanic origin: no Race: other race alone or in combination with white only	Hispanic origin: no Race: other race
Non-Hispanic white	Hispanic origin: no Race: white alone	Hispanic origin: no Race: white

Source: BBC Research & Consulting from the IPUMS program: <http://usa.ipums.org/usa/>.

Business ownership. BBC uses the Census “class of worker” variable (CLASSWKD) to determine self-employment. This variable is the same for 1980 and 2000 with one exception: the 1980 variable does not include a separate category for individuals who work for a wage or salary at a non-profit organization.

Changes in industry codes between censuses. The construction industry is coded as “077” for the 2000 Census and “060” for the 1980 Census. The 2000 Census represents the “architectural, engineering and related services” industry with code “729.” In 1980, the code is “882” for “engineering, architectural and surveying services.”

Changes in occupational codes between censuses. Occupational codes changed between 1980 and 2000 for the engineering industry. Coded as “53” in 1980, civil engineers were coded as “136” in 2000. Codes for occupations within the construction industry also changed between the two Censuses. Figure H-2 contains the occupational code crosswalk and all job descriptions.

Changes in educational variables between Censuses. The 1980 Census data now includes the same educational variable found in the 2000 Census data, although the questions used for each Census to capture educational attainment differed between the two surveys.⁵

⁵ For a more detailed explanation, see footnote 4.

Figure H-2.
Occupational crosswalk for 1980 and 2000 IPUMS data

Census 2000 occupational title and code	Census 1980 occupational title and code	Job description for 2000 titles
Construction managers 22	Managers and administrators, n.e.c. 19	Plan, direct, coordinate, or budget, usually through subordinate supervisory personnel, activities concerned with the construction and maintenance of structures, facilities, and systems. Participate in the conceptual development of a construction project and oversee its organization, scheduling, and implementation. Include specialized construction fields, such as carpentry or plumbing. Include general superintendents, project managers, and constructors who manage, coordinate, and supervise the construction process.
First-line supervisors/managers of construction trades and extraction workers 620	Supervisors (categories separated): brickmasons, stonemasons, and tile setters; carpenters and related workers; electricians and power transmission installers; painters, paperhangers and plasterers; plumbers, pipefitters and steamfitters; n.e.c.; and extractive occupations 553-558 & 613	Directly supervise and coordinate the activities of construction or extraction workers.
Construction laborers 626	Construction laborers 869	Perform tasks involving physical labor at building, highway, and heavy construction projects, tunnel and shaft excavations, and demolition sites. May operate hand and power tools of all types: air hammers, earth tampers, cement mixers, small mechanical hoists, surveying and measuring equipment, and a variety of other equipment and instruments. May clean and prepare sites, dig trenches, set braces to support the sides of excavations, erect scaffolding, clean up rubble and debris, and remove asbestos, lead, and other hazardous waste materials. May assist other craft workers. Exclude construction laborers who primarily assist a particular craft worker, and classify them under "helpers, construction trades."

2008 ACS data. BBC also used 2008 ACS data from IPUMS. Conducted by the U.S. Census Bureau, the ACS uses monthly samples to produce annually updated data for the same small areas as the 2000 Census long-form.⁶ Since 2005, the ACS has expanded to a roughly 1 percent sample of the population, based on a random sample of housing units in every county in the U.S. (along with the District of Columbia and Puerto Rico).

Applying the person-level population weights to the 3,000,657 observations included in the data, the 2008 ACS represents 304,059,728 people in the U.S. The 2008 ACS includes 36,353 observations for Oklahoma representing 3,642,361 individuals.

With the exception of a few minor differences, the variables available for the 2008 ACS dataset are the same as those available for the 2000 Census 5 percent sample.

Changes in race/ethnicity categories between the 2000 Census and 2008 ACS. The 2000 Census 5 percent sample and the 2008 ACS PUMS data use essentially the same numerical categories for the detailed race variable (RACED). However, in both the samples, any category that represented fewer than 10,000 people was combined with another category. As a result, some PUMS race/ethnicity categories that occur in one sample may not exist in the other, which could lead to inconsistencies between the two samples once the detailed race/ethnicity categories are grouped according to the seven broader categories. This issue is unlikely to affect all but a very small number of observations. PUMS categories that were available in 2000 but not 2008 (or vice versa) represented a very small percentage of the 2000 (or 2008) population. Categories for the Hispanic variable (HISPAN) remained consistent between the two datasets.

Other variables. Other variables used by BBC did not change between 2000 and 2008. The variables CLASSWKD, LABFORCE, IND, OCC, and EDUCD were consistent between datasets, with variable codes in each case representing the same categories.

Survey of Small Business Finances

The study team used the Survey of Small Business Finances (SSBF) to analyze the availability and characteristics of loans among small business enterprises.

The SSBF, conducted every five years by the Federal Reserve Board, collects financial data from non-governmental for-profit firms with fewer than 500 employees. This survey is a nationally representative sample, structured to allow for analysis of specific geographic regions, industry sectors, and gender and racial groups. The SSBF is unique in that it provides detailed data on both firm and owner financial characteristics. For the purposes of this report, BBC used the surveys from 1998 and 2003, which are available at the Federal Reserve Board website.⁷

Data for 1998. The 1998 SSBF includes information from 3,561 small businesses. The survey oversampled minority-owned businesses, allowing for a more precise analysis of how race and ethnicity may affect loan and financial outcomes.

⁶ U.S. Census Bureau. *Design and Methodology: American Community Survey*. Washington D.C.: U.S. Government Printing 2009. Available at http://www.census.gov/acs/www/SBasics/design_meth.htm

⁷ The Federal Reserve Board. *Survey of Small Business Finances, 1998* and *Survey of Small Business Finances, 2003*. Available online at <http://www.federalreserve.gov/pubs/>.

Categorizing owner race/ethnicity and gender. Definition of race and ethnic groups in the 1998 SSBF are slightly different than the classifications used in the 2000 Census and 2008 ACS. In the Survey, businesses are classified into the following five groups:

- Non-Hispanic white;
- Hispanic American;
- African American;
- Asian American;
- Native American; and
- Other (unspecified).

A business is considered Hispanic American-owned if more than 50 percent of the business was owned by Hispanic Americans, regardless of race. All businesses reporting 50 percent or less Hispanic American ownership are included in the racial group that owns more than half of the company. No firms reported ownership by “other.”

Similarly, firms were classified as female-owned if more than 50 percent of the firm was owned by female individuals. Firms owned half by female owners and half by male owners were included in the male category.

Defining selected industry sectors. In the 1998 SSBF, each business was classified according to SIC code and placed into one of eight industry categories:

- Construction;
- Mining;
- Transportation, communications and utilities;
- Finance, insurance and real estate;
- Trade;
- Engineering;
- Services (excluding engineering); or
- Agriculture, forestry and fishing.

Region variables. The SSBF divides the United States into nine regions. Oklahoma is included in the West South Central region.

Loan denial variables. In the 1998 survey, firm owners were asked if they have applied for a loan in the last three years and whether loan applications were always approved, always denied, or sometimes approved and sometimes denied. For the purposes of this study, only firms that were always denied were considered when analyzing loan denial.

Data for 2003. The 2003 SSBF differs from previous surveys in terms of the population surveyed, the variables available and in data reporting methodology.

Population differences. Similar to the 1998 survey, the 2003 survey records data from businesses with 500 or fewer employees. The sample contains data from 4,240 firms, but in 2003, minority-owned firms were not oversampled. In the 1998 data, 7.3 percent of the survey firms were owned by Hispanic Americans, but that number dropped to 4 percent in the 2003 data. Such percentages also dropped for African American-owned (7.7% to 2.8%) and Asian American-owned firms (5.7% to 4.2%). This aspect of the 2003 SSBF may affect the accuracy of analyses related to differences in loan application outcomes for race and ethnicity groups.

Variable differences. In the 2003 SSBF, businesses were able to give responses on owner characteristics for up to three different owners. The data also include a fourth variable that is a weighted average of other answers provided for each question. In order to define race/ethnicity and gender variables consistently from the 1998 to 2003 surveys, BBC used the final weighted average for variables on owner characteristics. Firms were then divided into race, ethnicity and gender groups according to the same guidelines used for the 1998 data.

Industry, region and loan denial variables for the 2003 survey were defined by the study team along the same guidelines as the 1998 survey with one exception. The 2003 survey did not include any firms in the agriculture, forestry and fishing industry.

Data reporting. Due to missing responses to survey questions in both the 1998 and 2003 datasets, data were imputed to fill in missing values. For the 1998 SSBF data, missing values were imputed using a randomized regression model to estimate values based on responses to other questions in the survey. A single variable includes both reported and imputed values, and a separate “shadow variable” can be used to identify where missing values have been imputed. However, the missing values in the 2003 data set were imputed using a different method than in previous studies. In the 1998 survey data, the number of observations in the data set matches the number of firms surveyed. However, the 2003 data includes five implicates, each with imputed values that have been filled in using a randomized regression model.⁸ Thus there are 21,200 observations in the 2003 data, five for each of the 4,240 firms surveyed. Across the five implicates, all reported values are identical, whereas imputed values may differ. In both data sets, therefore, when the firm provided an answer to a survey question, that value did not change during imputation; only missing values are predicted and filled in.

As discussed in a recent paper about the 2003 imputations by the Finance and Economics Discussion Series, missing survey values can lead to biased estimates and inaccurate variances and confidence intervals.⁹ These problems can be corrected through use of multiple implicates. In order to provide the most accurate analysis, BBC utilized all five implicates provided with the 2003 data in analysis of the survey.

⁸ For a more detailed explanation of imputation methods, see the “Technical Codebook” for the *2003 Survey of Small Business Finances*.

⁹ Lieu N. Hazelwood, Traci L. Mach and John D. Wolken. *Alternative Methods of Unit Nonresponse Weight Adjustments: An Application from the 2003 Survey of Small Businesses*. Finance and Economics Discussion Series Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board. Washington, D.C., 2007.
<http://www.federalreserve.gov/pubs/feds/2007/200710/200710pap.pdf>

Multiple implicates were not provided with the 1998 data, making method of analysis used for the 2003 data inapplicable. To address this, the study team performed analysis two different ways, first only with observations whose data was not imputed and second with all observations; differences in results were insignificant. For summary statistics using SSBF data, BBC included observations with missing values in the analyses. For the probit regression model presented in Appendix G, the study team did not include observations with imputed values for the depended variable, loan denial.

Survey of Business Owners (SBO)

BBC used data from the 2002 SBO to analyze mean annual firm receipts.

The SBO is conducted every five years by the U.S. Census Bureau. The most recent publication of the SBO was collected in 2002. (Data for the 2007 SBO are due to be released by the end of 2010). Response to the survey is mandatory, which ensures comprehensive economic and demographic information for business and business owners in the U.S. All tax-filing businesses are covered in the survey: firms with paid employees and firms without paid employees. In 2002, almost 23 million firms were surveyed.

BBC compiled a subset of the SBO data including number of firms, number of firms with paid employees, number of firms without paid employees, and total receipts. This information was available by geographic location, industry, gender and race/ethnicity.

Data are reported for the U.S. as well as for individual states.

The Census Bureau uses 2007 North American Industry Classification System (NAICS) for classification of industry. BBC analyzed data for firms in all industries and for firms in the construction industry.

To categorize business ownership of firms, the Census Bureau uses standard definitions for women-owned and minority-owned businesses. A business is defined as female-owned if more than half (51 percent or more) of the ownership and control is by women. Firms with joint male-/female-ownership were tabulated as an independent gender category. A business is defined as minority-owned if more than half (51 percent or more) of the ownership and control is African American, Asian, Hispanic, Native American or another minority group. Respondents had the option of selecting one or more racial groups when reporting on business ownership. BBC then analyzed data pertaining to the following six racial categorizations:

- African American
- Asian American
- Native American
- White
- Hispanic
- Non-Hispanic