Nonemployer Statistics by Demographics (NES-D): Using Administrative and Census Records Data in Business Statistics

by

Adela Luque
U.S. Census Bureau

Renuka Bhaskar
U.S. Census Bureau

James Noon
U.S. Census Bureau

Kevin Rinz
U.S. Census Bureau

Victoria Udalova
U.S. Census Bureau

CES 19-01 January, 2019

Abstract

The quinquennial Survey of Business Owners or SBO provided the only comprehensive source of information in the United States on employer and nonemployer businesses by the sex, race, ethnicity and veteran status of the business owners. The annual Nonemployer Statistics series (NES) provides establishment counts and receipts for nonemployers but contains no demographic information on the business owners. With the transition of the employer component of the SBO to the Annual Business Survey, the Nonemployer Statistics by Demographics series or NES-D represents the continuation of demographics estimates for nonemployer businesses. NES-D will leverage existing administrative and census records to assign demographic characteristics to the universe of approximately 24 million nonemployer businesses (as of 2015). Demographic characteristics include key demographics measured by the SBO (sex, race, Hispanic origin and veteran status) as well as other demographics (age, place of birth and citizenship status) collected but not imputed by the SBO if missing. A spectrum of administrative and census data sources will provide the nonemployer universe and demographics information. Specifically, the nonemployer universe originates in the Business Register; the Census Numident will provide sex, age, place of birth and citizenship status; race and Hispanic origin information will be obtained from multiple years of the decennial census and the American Community Survey; and the Department of Veteran Affairs will provide administrative records data on veteran status.

The use of blended data in this manner will make possible the production of NES-D, an annual series that will become the only source of detailed and comprehensive statistics on the scope, nature and activities of U.S. businesses with no paid employment by the demographic characteristics of the business owner. Using the 2015 vintage of nonemployers, initial results indicate that demographic information is available for the overwhelming majority of the universe of nonemployers. For instance, information on sex, age, place of birth and citizenship status is available for over 95 percent of the 24 million nonemployers while race and Hispanic origin are available for about 90 percent of them. These results exclude owners of C-corporations, which represent only 2 percent of nonemployer firms. Among other things, future work will entail imputation of missing demographics information (including that of C-corporations), testing the longitudinal consistency of the estimates, and expanding the set of characteristics beyond the demographics mentioned above. Without added respondent burden and at lower imputation rates and costs, NES-D will meet the needs of stakeholders as well as the economy as a whole by providing reliable estimates at a higher frequency (annual vs. every 5 years) and with a more timely dissemination schedule than the SBO.

Keyword*:* Nonemployer, administrative records, census records, blended data, demographics, business statistics, business owners, NES-D, NES, SBO, ABS.

---

[*] Acknowledgements: This work is the result of a collaborative effort involving multiple Divisions in the Economic and Research & Methodology Directorates at the U.S. Census Bureau. Hence, we want to thank all members of the governance, advisory and working groups who are part of this project.

Corresponding author: Adela.luque@census.gov

Disclaimer: All results have been reviewed to ensure that no confidential information is disclosed. The statistical summaries reported in this paper have been reviewed and approved by the Census Bureau's Disclosure Review Board (DRB items #CBDRB-FY18-449 and #CBDRB-FY19-075).

Executive Summary

Background

The costs of household and business surveys have been increasing as response rates have declined, while at the same time, the public and stakeholders across the economy require reliable, and more frequent and timely data. In response to these trends, the U.S. Census Bureau consolidated three key business surveys into one new program, the Annual Business Survey or ABS.[1] Two of the consolidated surveys, the SBO and ASE, were the official sources of demographic information on employer and nonemployer businesses and their owners by race, ethnicity, sex, and veteran status. The SBO provided demographics estimates for both nonemployer and employer firms while the ASE covered only employer firms.

The consolidation transferred the employer component of the SBO to the ABS; however, the ABS does not survey nonemployer businesses.[2] As a result, the nonemployer component of the SBO will now be accomplished through a new blended-data approach that leverages existing administrative records (AR) and census records to assign demographic characteristics to the universe of approximately 24 million nonemployer firms.[3] By using administrative records, Census will be able to produce without added respondent burden and at lower costs an annual series that will become the sole source of detailed and comprehensive statistics on the scope, nature and activities of U.S. businesses with no paid employment by the demographic characteristics of the

[1] The consolidated surveys are: the Survey of Business Owners (SBO), the Annual Survey of Entrepreneurs (ASE) and the Business R&D and Innovation Survey for Microbusinesses (BRDI-M). See Appendix 1 for a general description of these surveys and Foster and Norman (2018) for more details on the ASE.
[2] The ABS will provide annual estimates of demographics for employer businesses, and thus on a more frequent basis than the quinquennial SBO.
[3] As of 2015.

business owners.[4] The new series is called the Nonemployer Statistics by Demographics series or NES-D.

Nonemployer Statistics by Demographics Series Content

The nonemployer universe is comprised of businesses with no paid employment or payroll, with annual receipts of $1,000 or more ($1 or more in the construction industries), and filing tax forms for sole proprietorships (Form 1040, Schedule C), partnerships (Form 1065), or corporations (the Form 1120 series). The vast majority of nonemployers are sole proprietors. As of 2015, 86 percent of nonemployer firms were sole proprietorships, 7 percent were partnerships, 5 percent were S-corps, and the remaining 2 percent were C-corporations. However, partnerships and S-corporations over-account for total nonemployer receipts. Specifically, partnerships make up 22 percent, and S-corps 11 percent of total receipts.

Without NES-D, the official estimates of the demographics of U.S. nonemployer businesses previously provided by the SBO would cease to exist. NES-D will consist of summary statistics of number of establishments, business owners,[5] and receipts of U.S. nonemployer businesses by the demographic characteristics of the business owner as well as by the legal form of organization and receipts-size class of the business at detailed industry and geography levels. In this way, NES-D also represents the expansion of the content of the existing annual Nonemployer Statistics series (NES) by adding the demographics dimension to the NES.[6]

---

[4] The annual Nonemployer Statistics series (NES) provides establishment counts and receipts for nonemployers but contains no demographic information on the business owners.

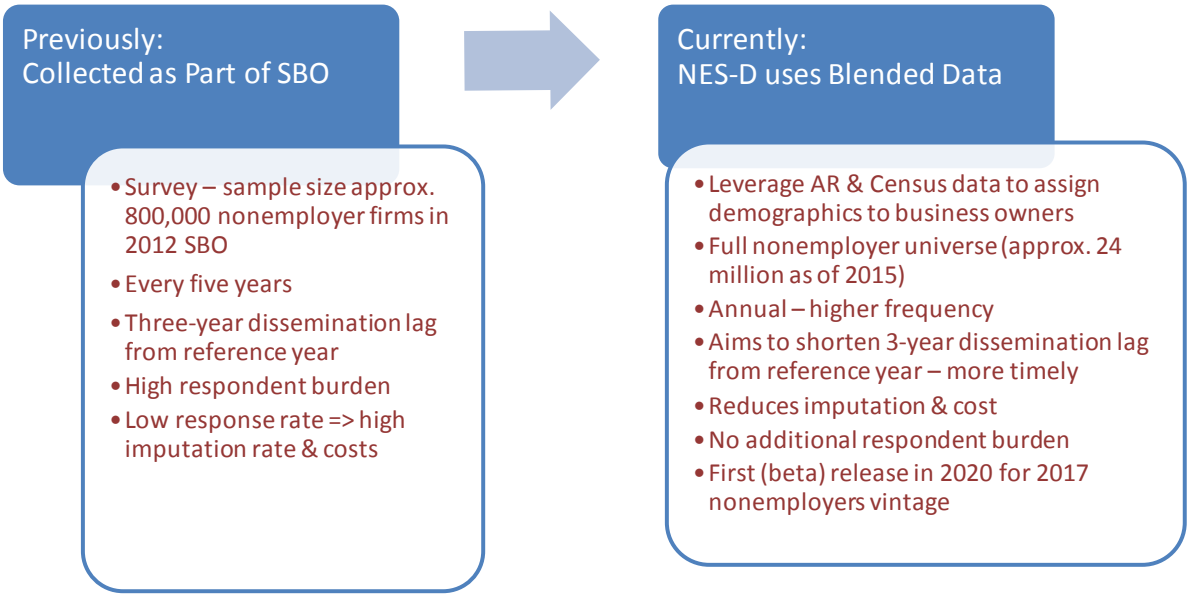
[5] Business owner counts will not include owners of C-corporations since AR data do not contain information on owners of this type of firms. C-corporations are further discussed in the Methodology section of this paper.

[6] See https://www.census.gov/programs-surveys/nonemployer-statistics/about.html for a description of NES.

NES-D will include key demographic characteristics (i.e., sex, race, Hispanic origin, and veteran status)[7] that were collected by the SBO and imputed if missing, as well as demographics that the SBO collected but did not impute if missing (i.e., age, place of birth, and citizenship). The demographic characteristics as well as the universe of nonemployer businesses itself come from a spectrum of administrative records and census data sources including the Business Register (BR), tax data, the Decennial Census and American Community Survey (ACS), Census Numident files, and administrative records on veteran status from the Department of Veteran Affairs (VA). Future work versions of NES-D will expand in content to include additional characteristics that could improve our understanding of nonemployers dynamics. Examples may include: household attributes obtainable through tax Form 1040 (such as marital status, number of dependents or home ownership), transitions from nonemployer to employer status, or information on whether nonemployers' income is the primary source of income for the nonemployer business owner, which can be obtained through W-2 tax data.

The figure below contrasts how the SBO produced demographic estimates for nonemployer businesses vs. the blended-data NES-D approach.

[7] These characteristics were referred to as "core demographics" in the SBO.



The SBO was conducted every five years and disseminated estimates with a 3 year lag (from the reference year). Collecting information from nonemployers was harder and more costly than it was from employers, and this is reflected in higher non-response rates for nonemployer firms. For instance, even though the law mandated survey participation in the SBO, only about 65 percent of the mailed cases returned the questionnaire, and item non-response rates for nonemployers could be as high as approximately 50 percent - while for employer firms it was about 25 percent.[8]

The right side of the figure shows NES-D, which leverages administrative records and previous census records data (i.e., decennial and ACS) to assign demographics to the entire universe of nonemployers to produce an annual series of statistics with no additional respondent burden, and at lower costs and imputation rates than previously produced via survey. Our goal is to eventually have a dissemination lag shorter than the SBO's 3 years (from the reference year). Also, we hope that the combination of the size of the nonemployer universe together with NES-D's lower

[8] See https://www2.census.gov/programs-surveys/sbo/methodology/2007/sbo_nonresponse_analysis.pdf.

imputation rates will allow the publication of cell counts at detailed levels of geography and industry for small demographic groups, which the SBO had to suppress because of data quality and/or disclosure concerns. On the other hand, disclosure avoidance rules are evolving and often becoming more stringent, which may inhibit this effort. Some information provided by the SBO (e.g., sources of business funding or language used in business transactions) are not available in AR, so there will be some loss of information relative to the SBO. However, because NES-D will produce high quality, more frequent and timely data with no respondent burden, stakeholders and data users find the tradeoff is worth it.

Initial Steps and Preliminary Results

This paper describes the initial stages of creating a NES-D prototype, and also provides some preliminary tabulations. The first step in the creation of NES-D consists in identifying the nonemployer universe and extracts it from the BR[9]. The BR is a comprehensive database of all U.S. employer and nonemployer business establishments developed and maintained by the U.S. Census Bureau, with data spanning from 1975 to the present. It provides information on receipts, industry, and the geographic location of the business. An essential piece is the assignment of anonymized personal identifiers to individuals in AR, tax and census data sources upon data arrival at the Census Bureau.[10] These personal identifiers are called Protected Identification Keys or PIKs. The BR already contains PIKs for sole proprietors from tax Form 1040, while we obtain PIKs for owners of partnerships and S-corporations from Schedule K-1 tax data.[11] PIKs are then used as

[9] This is done by Census' Economic Directorate.
[10] See Wagner & Layne (2014) for more information on Census' probabilistic algorithm that assigns anonymized individual identifiers to individual data sources including decennial and ACS, other survey data as well as tax and other AR data.
[11] A detailed explanation of this step is included in the Methodology section of this paper.

linking keys across data sources to obtain information on the demographic characteristics of the business owners. Owners of C-corporations, though, cannot be unequivocally identified through tax or other administrative data sources; thus, assignment of PIKs for this group of owners is not possible. Our goal is to impute demographic characteristics for these firms.[12] Fortunately, C-corporations account for only 2 percent of the nonemployer universe and 4 percent of total nonemployer receipts.

Our initial owner-level results, based on the 2015 nonemployer file, indicate that demographic information is available for the overwhelming majority of the nonemployer population. PIK information is available for over 95 percent of all nonemployer businesses, and match rates to AR and census data sources are also very high. Specifically, matching to the Census Numident provides sex, age, place of birth and citizenship status for approximately 99 percent of owners with PIKs while Decennial and ACS data supply race and Hispanic origin for approximately 90 percent of owners with PIKs. In fact, about 90 percent of owners with PIKs have no missing demographics and only about 1 percent is missing three or more demographic characteristics.[13] By contrast, even though the law mandated survey participation in the SBO, only approximately 65 percent of the mailed cases returned the questionnaire and unit non-response rates (blank responses to individual questions) for nonemployers was approximately 50 percent.

While the assignment of demographic characteristics to identifiable business owners is straightforward, the aggregation of this information to the firm level can be more complex for businesses with more than one owner (i.e., partnerships and S-corporations). As in the SBO, we

[12] See a discussion of this topic in the Methodology section of this paper.
[13] As already noted, these PIK coverage results do not include owners of C-corporations.

assign firms to demographic groups by determining the total share of firm ownership held by individual members of each demographic group. A business is assigned to a given demographic group if the group's owners account for a majority stake (more than 50 percent) in the firm. We currently follow SBO's methodology and consider only the four owners with the largest ownership shares in the business, and only firms where the largest owner owns at least 10 percent of the business. While some firms have more than 4 owners, we do not consider this restriction a major source of noise since over 90 percent of partnerships and about 98 percent of S-corporations have four owners or less. In addition, it is not conceptually obvious that it makes sense to (demographically) categorize firms with diffuse ownership (i.e., where there is no owner with at least 10 percent ownership).[14]

In this paper, we calculate preliminary firm-level estimates of nonemployer demographics, but these should be interpreted with caution since they do not include imputed values of missing demographics, and are incomplete. Therefore, at this early stage, they are not intended to be representative of the demographics of the underlying nonemployer population. Although these preliminary results are not fully comparable to prior SBO publications, we nevertheless undertake an initial comparison to see if they behave according to our expectations, and overall they do – as discussed below.

Starting with race, our results indicate that the AR-based race distribution and the SBO distributions (for 2007 and 2012) are within 5 percentage points, with the largest difference found in white-owned businesses (see Table 30). Specifically, approximately 81 percent of 2015

[14] We also conduct sensitivity analysis by relaxing the four owner and 10 percent rules in the paper. We find that relaxing the four-owner rule has virtually no impact and that relaxing the 10 percent rule increases firm demographic assignment by approximately 2.5 percent.

nonemployer businesses are white-owned according to our preliminary AR-based estimates while the 2007 SBO reported 85 percent, and the 2012 SBO 76 percent. The AR and SBO estimate differences for the other race categories are approximately within 2 percentage points. A few precautionary notes are in order when looking at these race results though: i) Our estimates are calculated excluding the approximately 10 percent of owners with missing race, ii) the 2007 and 2012 SBOs differed in the way they categorize the race of individuals that entered a Hispanic or Latino response in the race write-in boxes,[15] iii) the SBO included a "Some-Other-Race" category while we do not,[16] iii) we include a "Multiple-Race" category while the 2012 SBO did not.[17] In the SBO, individuals of multiple races were assigned into their corresponding race categories. For instance, an owner who reported to be both Asian and White was counted separately as Asian and White in the SBO tabulations. As a result, in the SBO, businesses could be tabulated in more than one racial group because i) a sole owner reported to be of more than one race, ii) a majority owner reported to be of more than one race, or iii) a majority combination of owners was reported to be of more than one race. Our future estimates will include imputation of missing values, and the assignment of individuals of multiple races to their corresponding race categories just like in the SBO.

Regarding Hispanic origin results (Table 31), we see that the difference between the AR-based distribution and the 2007 and 2012 SBOs are within 3 percentage points. Approximately 12 percent

[15] In the 2012 SBO, if a respondent entered a Hispanic or Latino ethnicity in the race write-in box, the record was categorized as "Some-Other-Race". By contrast, in the 2007 SBO, that same case would have been categorized as "White". The change was implemented to be consistent with 2010 Census methodology.
[16] In order to adhere to the revisions to Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting issued by the Office of Management and Budget, NES-D will not allow for a race category of "Some Other Race".
[17] Approximately 2 percent of owners were of multiple races.

of nonemployer businesses are Hispanic-owned according to our preliminary AR-based results while the 2007 and 2012 estimates were 9 and 14 percent respectively. Again, note that our calculation excludes individuals with missing Hispanic origin information.

Regarding results on sex categories, and as expected, the AR-based percentage of nonemployer businesses equally owned by men and women is notably lower than the percentage obtained from the SBO (see Table 18). This is because the SBO allowed single-owner firms to enter a response indicating that the business was owned equally by a man and a women (usually married couples) – even when the business was officially owned by only one person.[18] By contrast, following the sex of the single owner identified on the sole proprietor tax Form 1040, AR data only allow us to classify sole proprietorships as either male or female-owned. Since the vast majority of nonemployer businesses (86 percent in 2015) are sole proprietorships, the AR-based percentage of the equally owned category is considerably lower than the percentage obtained from the SBO. Specifically, our preliminary results on nonemployer businesses equally owned by men and women are 6 percentage points lower from the 2012 SBO and more than 10 percentage points lower from the 2007 SBO estimates.[19]

Our preliminary AR-based estimate of veteran-owned businesses indicates 6 percent of nonemployer businesses were veteran-owned in 2015, while prior SBOs estimated that figure to be 9 percent for 2007 and 2012 nonemployers (see Table 35). Again, the lower AR-based estimate was

[18] There are some exceptions. Married couples can legally jointly own a sole proprietorship if they file taxes as a "qualified joint venture".

[19] The equally-owned percentage is lower in the 2012 SBO because in 2012 the SBO used administrative records data to direct replace or impute (whenever direct replacement was not possible) sex, race, Hispanic origin and veteran status for nonemployer sole proprietors. This partially explains the large decrease in the percentage of equally-owned businesses by men and women from the 2007 to the 2012 SBO.

expected since the concept of veteran captured in the SBO is broader than what the VA identifies as a veteran, and according to our own analysis, older veterans are under-represented in VA's AR data.[20] Additional AR data source - the Department of Defense's Defense Enrollment Eligibility Reporting System (DEERS) database - may allow us to complement the USVETS data to better align the SBO/ABS' definition of a veteran with the one we can obtain using AR data. Future work will include the examination of this possibility.

Conclusions and Next Steps

The creation of NES-D illustrates how valuable leveraging existing individual-level AR and census records can be in creating business statistics. Without incurring additional respondent burden and with substantially reduced costs, NES-D will produce reliable, and more frequent and timely estimates of nonemployers demographics than the survey it replaces.

NES-D represents an innovative approach to producing business statistics whose methodology is also well grounded in a body of proven administrative records research. This research provides evidence of the suitability of the demographic data sources employed in NES-D to direct replace demographic information in household and business surveys, and sheds light into non-sampling errors underlying those data (e.g., coverage issues, conceptual and timing misalignments, biases in PIK assignment or misreporting). [21] Many of these issues primarily apply to hard-to-count populations, who often are not well represented in tax data.[22] Fortunately, NES-D's nonemployer

[20] A full discussion of this analysis is included in the Data and Appendix sections of the paper.
[21] See, for instance, Bhaskar (2016), Ennis (2016), Luque (2016), Noon (2016), Rastogi & O'Hara (2012), Bhaskar et al. (2014), Luque & Bhaskar (2014), Bond et al. (2014). Also, a discussion of these issues is included in the Challenges section of this paper.
[22] Certain populations are missed at higher rates and are under-represented in decennial data. These are referred to as "hard-to-count" populations and include very young children, racial and ethnic minorities, low income persons, immigrants not yet fully integrated in the economy, people in rural communities and mobile persons.

universe is well represented in tax data, and therefore, not as impacted by these concerns relative to the general U.S. population.

Other challenges pertain to matters related to data acquisition and specific use, as well as limitations due to disclosure avoidance rules. For instance, data use agreements between Census and the various government agencies owning the AR data sources are essential. It is also critical that the AR data sources are consistently available over time, without substantial changes in format, and that the data are delivered in a timely fashion and at the frequency necessitated by NES-D. We will work to meet these challenges, and provide clarity and transparency. Stakeholders in particular, and economic agents in general, are more reliant on data than ever before. To be useful, these data have to be accurate, timely, frequent, consistent, credible and transparent. It is our goal to work to have NES-D fulfill these criteria.

NES-D is in its nascent stage, and although there are challenges along the way, the initial findings are very promising. As our results show, demographic information can be found in AR data for the overwhelming majority of nonemployers. During the next year we plan to test the longitudinal consistency of our estimates, produce count and receipts estimates by geographic and industry detail, and address imputation of missing demographics - including the viability of imputing demographics for C-corporations, and potential improvements to current imputation methodology. In the third year, this work will transition into the production phase. The goal is to release a beta version of NES-D in 2020 with the 2017 nonemployer vintage.

Nonemployer Statistics by Demographics (NES-D)

Using Administrative and Census Records Data in Business Statistics

Adela Luque, Renuka Bhaskar, James Noon, Kevin Rinz, Victoria Udalova

U.S. Census Bureau

Abstract

The quinquennial Survey of Business Owners or SBO provided the only comprehensive source of information in the United States on employer and nonemployer businesses by the sex, race, ethnicity and veteran status of the business owners. The annual Nonemployer Statistics series (NES) provides establishment counts and receipts for nonemployers but contains no demographic information on the business owners. With the transition of the employer component of the SBO to the Annual Business Survey, the Nonemployer Statistics by Demographics series or NES-D represents the continuation of demographics estimates for nonemployer businesses. NES-D will leverage existing administrative and census records to assign demographic characteristics to the universe of approximately 24 million nonemployer businesses (as of 2015). Demographic characteristics include key demographics measured by the SBO (sex, race, Hispanic origin and veteran status) as well as other demographics (age, place of birth and citizenship status) collected but not imputed by the SBO if missing. A spectrum of administrative and census data sources will provide the nonemployer universe and demographics information. Specifically, the nonemployer universe originates in the Business Register; the Census Numident will provide sex, age, place of birth and citizenship status; race and Hispanic origin information will be obtained from multiple years of the decennial census and the American Community Survey; and the Department of Veteran Affairs will provide administrative records data on veteran status.

The use of blended data in this manner will make possible the production of NES-D, an annual series that will become the only source of detailed and comprehensive statistics on the scope, nature and activities of U.S. businesses with no paid employment by the demographic characteristics of the business owner. Using the 2015 vintage of nonemployers, initial results indicate that demographic information is available for the overwhelming majority of the universe of nonemployers. For instance, information on sex, age, place of birth and citizenship status is available for over 95 percent of the 24 million nonemployers while race and Hispanic origin are available for about 90 percent of them. These results exclude owners of C-corporations, which represent only 2 percent of nonemployer firms. Among other things, future work will entail imputation of missing demographics information (including that of C-corporations), testing the longitudinal consistency of the estimates, and expanding the set of characteristics beyond the demographics mentioned above. Without added respondent burden and at lower imputation rates and costs, NES-D will meet the needs of stakeholders as well as the economy as a whole by providing reliable estimates at a higher frequency (annual vs. every 5 years) and with a more timely dissemination schedule than the SBO.

Keywords: Nonemployer, administrative records, census records, blended data, demographics, business statistics, business owners, NES-D, NES, SBO, ABS.

I. Introduction

Like their household counterparts, business surveys have endured declining response rates and increasing costs. In an effort to address these issues while maintaining data quality, reducing respondent burden and improving timeliness, frequency and efficiency, three business surveys have been consolidated into one new survey, the Annual Business Survey or ABS. The consolidated surveys are the five-year Survey of Business Owners (SBO), the Annual Survey of Entrepreneurs (ASE) and the Business R&D and Innovation Survey for Microbusinesses (BRDI-M).[23] The SBO and ASE were the official sources of information on selected economic and demographic characteristics for employer and nonemployer businesses and their owners by sex, ethnicity, race, and veteran status. The SBO, in particular, provided demographics estimates for both nonemployer and employer firms while the ASE focused just on employer firms.

While the new ABS will survey employer firms, it will not cover nonemployer businesses.[24] The continuation of estimates of demographics for nonemployer businesses (previously undertaken by the SBO) will now be accomplished by leveraging existing administrative records (AR) and census records to assign demographic characteristics to the universe of approximately 24 million nonemployer firms.[25] The use of administrative records in this manner will make possible the production of an annual series that will become the only source of detailed and comprehensive statistics on the scope, nature and activities of U.S. businesses with no paid employment by the

[23] See Appendix 1 for a general description of these surveys, and Foster and Norman (2018) for more details on the ASE.
[24] The ABS will provide annual estimates of demographics by employer businesses, and thus on a more frequent basis than the quinquennial SBO.
[25] As of 2015.

demographic characteristics of the business owner. Demographics will include characteristics collected by the SBO that were imputed if missing (i.e., sex, race, Hispanic origin, and veteran status)[26] as well as demographics that the SBO collected but did not impute if missing (i.e., age, place of birth, and the citizenship of the business owner).

We will refer to this new series as the Nonemployer Statistics by Demographics series or NES-D from this point forward. NES-D will then consist of summary statistics of number of establishments, business owners, and receipts of U.S. nonemployer businesses by the demographic characteristics of the business owner as well as by the legal form of organization and receipts-size class of the business at detailed industry and geography levels.[27] In this way, NES-D also represents the expansion of the content of the existing annual Nonemployer Statistics series (NES) by adding the demographics dimension to the NES.[28] Without NES-D, the official estimates of the demographics of U.S. nonemployer businesses previously provided by the SBO would cease to exist.

In today's economy, stakeholders and the public as a whole require access to comprehensive, accurate, timely, frequent and consistent data– and NES-D aims to fulfill that demand. It will reduce respondent burden for the eligible nonemployer population, which has implications for increasing non-response rates. Even though the law mandated survey participation in the SBO, only approximately 65 percent of the mailed cases returned the questionnaire – which is low relative to other mandatory federal surveys. Furthermore, the non-response rates for nonemployers were higher than for employers. Specifically, item non-response rates were approximately 50 percent for

[26] These characteristics were referred to as "core demographics" in the SBO.
[27] Business owner counts will not include owners of C-corporations since AR data do not contain information on owners of this type of firms. C-corporations are further discussed in the Methodology section of the paper.
[28] See https://www.census.gov/programs-surveys/nonemployer-statistics/about.html for a description of NES.

nonemployers and 25 percent for employer firms in the 2007 SBO, with some demographic groups experiencing lower response rates than others.[29] This issue not only impacted survey costs but also potentially the quality of the estimates since a large share of the demographic information had to be imputed.[30] Also, because NES-D will use the entire universe of nonemployers and quality AR demographic data with low rates of missingness, it will likely be able to show counts for smaller demographic groups at detailed geography and industry levels[31] - counts that previously had to be suppressed due to disclosure rules or data quality (i.e., high relative standard erros). Finally and most importantly, the series will provide more timely and frequent estimates than the legacy SBO. While the SBO was conducted every five years with a dissemination lag of three years (from the reference year), NES-D will be available on an annual basis, and ultimately be available with a shorter dissemination schedule. We should note though that because NES-D employs AR data, some of the information the SBO collected on businesses (e.g., language used in business transactions or sources of business funding) may not be obtainable through AR.[32] At the same time, one of the goals is to expand the content of NES-D in the future to include additional characteristics that will improve our understanding of nonemployers dynamics. Examples may include: household attributes such as marital status, number of dependents or home ownership, transitions from nonemployer to employer status, or information on whether nonemployers' income is the primary source of income for the nonemployer business owner.

[29] See https://www2.census.gov/programs-surveys/sbo/methodology/2007/sbo_nonresponse_analysis.pdf.
[30] See Fricker & Tourangeau (2010) for evidence of how higher non-response rates lead to lower quality estimates, and the Challeges section of this paper for a discussion on non-sampling errors in AR and census records data.
[31] This is subject to disclosure avoidance rules.
[32] See https://www.census.gov/data/tables/2012/econ/sbo/2012-sbo-characteristics-of-businesses.html for business characteristics tables provided by the SBO.

NES-D represents an innovative approach to producing business statistics, but importantly, the construct is well grounded in prior substantive administrative records research. This research provides evidence of the suitability and quality of the data sources employed in NES-D to direct replace demographic information in household and business surveys.[33] Thanks to this research we have a good understanding of potential non-sampling errors in our sources of demographic data, such as coverage issues, conceptual misalignments, biases in PIK assignment or misreporting. Many of these issues primarily apply to hard-to-count populations, who often are not well represented in tax data.[34] Fortunately, the nonemployer business owner universe is well represented in tax data and is not as impacted by these concerns as the general U.S. population. We discuss these issues in the Challenges section of the paper.

This paper describes the initial stages of creating a NES-D prototype and provides preliminary tabulations of the demographic characteristics of businesses and business owners for all industries at the national level. Because of their preliminary nature, these tabulations are not intended to be representative of the underlying nonemployer universe demographics. In addition, the current results do not include C-corporations since owners of such companies cannot be identified using tax or other AR data. This issue is discussed in more detail in the Methodology section of the paper.[35] Assigning demographics to that group of firms will be addressed in future work. For this paper we use the 2015 reference year of nonemployers, which was the latest nonemployer data

[33] See, for instance, Bhaskar (2016), Ennis (2016), Luque (2016), Noon (2016), Rastogi & O'Hara (2012), Bhaskar et al. (2014), and Bond et al. (2014).

[34] Certain populations are under-represented in decennial data. These are referred to as "hard-to-count" populations and include very young children, racial and ethnic minorities, low income persons, immigrants not yet fully integrated in the economy, people in rural communities and mobile persons.

[35] C-corporations are corporations in which the owners, or shareholders, are taxed separately from the entity. The taxing of profits from the business is at both corporate and personal levels, creating a double taxation situation. C-corporations are discussed further below.

available at the time this work began. In 2019 we will examine the longitudinal consistency of the estimates, provide receipts estimates, explore estimates by geographic and industry detail, and address imputation of missing demographics - including the viability of imputing demographics for C-corporations and potential improvements to current imputation methodology. In the third year, this work will transition into the production phase. The goal is to release a beta version of NES-D in 2020 with the 2017 nonemployer vintage.[36]

This paper is organized as follows. In the next section, we describe the data used in our analysis. Section III describes how we link the different AR and census data sources, and discusses methodological issues related to the assignment of demographics at the individual and firm levels. In Section IV we present results on the share of nonemployer business owners that are "linkable" to demographic AR and Census sources. Section V presents results on the availability of demographic characteristics for our nonemployer universe, and the prevalence of "missingness" across data sources. In Section VI we present and discuss initial tabulations of nonemployer demographics at the individual as well as firm level. Please note that all values in tables and figures in the paper have been rounded to four significant digits as part of Census' disclosure avoidance protocol. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000 and above 10,000,000 to the nearest 10,000. In Section VII we discuss challenges and limitations, and finally conclude and go over next steps in Section VIII.

[36] Pending all necessary reviews.

II. Data

NES-D is created from a variety of administrative records and census data sources that include the Business Register (BR), tax data, Decennial Census and American Community Survey (ACS) data, Social Security Administration (SSA) Numident data, and AR on veteran status.

The Business Register is a comprehensive database of all U.S. employer and nonemployer business establishments developed and maintained by the U.S. Census Bureau, with data beginning in 1975 and to the present. The primary source of BR data is annual or quarterly business payroll and income tax filings with the IRS. These data contain business name, identifier, address, industry classification, legal form of organization, receipts, and employment and payroll (these last two apply only to employer firms).

The universe of nonemployer businesses used in this work is identified and extracted from the BR by the Economic Directorate at the Census Bureau.[37] Specifically, the nonemployer universe is comprised of businesses with no paid employment or payroll, with annual receipts of $1,000 or more ($1 or more in the construction industries), and filing tax forms for sole proprietorships (Form 1040, Schedule C), partnerships (Form 1065), or corporations (the Form 1120 series). In the case of nonemployer sole proprietorships, the business identifier coincides with that of its owner – it is simply the Social Security Number (SSN) of the business owner obtained from tax Form 1040.[38] For

[37] See https://www.census.gov/programs-surveys/nonemployer-statistics/technical-documentation/methodology.html for additional details on how the Census Bureau identifies the nonemployer universe.
[38] SSNs are anonymized upon arrival at the Census Bureau and converted into Protected Identification Keys or PIKs, which are discussed below.

nonemployer partnerships and corporations, the business identifier is the Employer Identification Number or EIN.

Sole proprietorships are businesses owned and managed by one individual.[39] The owner or sole proprietor does not pay separate income tax on the company, but instead reports all losses/profits from the business on his/her individual IRS 1040 tax return. A partnership is a business with two or more owners, each receiving a share of the profits/losses of the business. A partnership must file an annual information return (Schedule K-1) to report the income/losses from its operations, but it is not subject to income tax itself. Instead, it "passes through" any profits/losses to its partners (hence their “pass-through entities” alias). Each partner includes his or her share of the partnership's income/loss on his/her personal tax return. S-corporations (or S-corps from this point forward) are corporations owned by one or more individuals (up to 75). Just like partnerships, they are also pass-through entities in that they pass the business’ corporate income/losses onto their shareholders for federal tax purposes, who then report their share of it on their personal tax returns. C-corporations (or C-corps from this point forward) are corporations in which the owners or shareholders, are taxed separately from the entity. The corporation itself is also subject to corporate income taxation. The taxing of profits from the business is at both corporate and personal levels, creating a double taxation situation. Only C-corps that are privately-owned are eligible for assignment of demographics (although we will refer to them simply as C-corps in this paper).[40] The challenge we face in regards to C-corps is that there is no tax Form or business

[39] One exception to this one-owner rule is that married couples can own sole proprietorships jointly as “qualified joint ventures”. Spouses who choose to have their businesses organized in this way must both participate in the business and file a joint income tax return, and the business may not have any members except for the couple. Another exception is “common property states”.

[40] Publicly-held C-corps, by definition, do not have identifiable owners.

registry that clearly identifies their owners. For this reason, C-corps are not included in the results presented in this report. In the next section, we discuss how we plan to address C-corps in the future.

Fortunately, C-corps constitute only 2 percent of the nonemployer universe and 4 percent of receipts. The vast majority of nonemployers are sole proprietors. As *Figure 1* shows, as of 2015, 86 percent of nonemployers fell under this category, 7 percent were partnerships, 5 percent were S-corps, and the remaining 2 percent were C-corporations. Partnerships and S-corps over-account for total nonemployers receipts. Specifically, partnerships account for 22 percent and S-corps for 11 percent. All North American Industry Classification System (NAICS) industries are included in the nonemployer database with some exceptions: crop and animal production, investment funds, trusts, and other financial vehicles, management of companies and enterprises, and public administration.[41]

To attach demographic characteristics to nonemployer business owners, we use anonymized unique individual identifiers that the Census Bureau assigns to individuals in AR and census data sources upon data arrival at the Bureau. These individual identifiers, which are known as the Census Bureau's Protected Identification Keys or PIKs, are used as linking keys to obtain demographic information from data sources, and attach those demographic characteristics to owners of nonemployer businesses. PIKs are assigned through the Census Bureau's Person Identification Validation System, which uses probabilistic record linkage techniques.[42] Census

[41] There are also certain industries are automatically reclassified when they appear in the nonemployer universe because these industries are generally assumed to require employees. For a list of these industries, see https://www.census.gov/programs-surveys/nonemployer-statistics/technical-documentation/methodology.html#par_textimage_36648475.

[42] See Wagner and Layne (2014) for more information.

employs tax data to assign PIKs to business owners. Depending on the legal form of organization (LFO) of the business, two IRS forms are used in the assignment: IRS Form 1040 for sole proprietors, and Schedule K-1 for owners of partnerships and S-corps. Owners of sole proprietorships report income from those businesses and identify the particular businesses from which it is derived when they file Form 1040 with their income taxes. This tax information, including the PIK of the owner, is already included in the BR, and hence, in our nonemployer database. When 1040 data arrives at Census, SSNs are anonymized and converted to PIKs. For partnerships and S-corps, we obtain owner PIKs from Schedule K-1. As described above, partnerships and S-corps prepare a Schedule K-1 to report each owner's share of income or loss. K-1s are provided to the IRS with the partnership or S-corp's tax return, and also to each owner so that they can add the information to their own tax returns (Form 1065 in the case of partnerships, and Form 1120S in the case of S-corps). In this way, Schedule K-1 data contain the EIN identifying the business itself (i.e., partnership or S-corp) as well as the owners of the business. Specifically, K-1 records contain the firm's EIN along with the SSNs of the owners (which are anonymized and converted into PIKs at their arrival at the Census Bureau).[43] Hence, for a given partnership or S-corp we have EIN-PIK pairs, which identify the firm and all its owners. Since our nonemployer database contains EINs, we are able to identify which partnerships and S-corps are nonemployers in K-1 data. The K-1 record also includes the share of the business owned by each owner. As we will see in the Methodology section, this information will be important in assigning demographic characteristics at the firm level.

[43] Partnerships and S-corps have to obtain EINs regardless of whether they have any employees.

Several data sources provide the demographic characteristics that NES-D will employ and are described below.

Previous Census Records

For the assignment of race and Hispanic origin to nonemployer business owners, we use previous decennial Census and American Community Survey (ACS) records. In this paper we specifically use the 2000 and 2010 decennial and ACS records from 2001-2015. This file, henceforth referred to as the 2015 previous census records or PCR file, uses the most recent data from previous census records to assign race and Hispanic origin responses; therefore, priority is given to 2011-2015 ACS data, then the 2010 Census, followed by 2001-2010 ACS data, and finally Census 2000. We are also evaluating the potential use of an Administrative Records Composite file for assigning race and ethnicity. Appendix 2 contains an extensive discussion of that file.

Census Numident

The SSA Numident file contains all transactions ever recorded against any single SSN - with each entry representing an addition or change (such as name changes) to the SSN record. This file is edited at the Census Bureau to create the Census Numident, which contains one record for each anonymized SSN or PIK.[44] Each PIK record in the Census Numident contains name, date of birth, sex, place of birth, citizenship status,[45] and date of death. In our work, the Census Numident provides the sex, age, place of birth and citizenship status of the business owner. Place of birth

[44] Name edits, DOB reconciliation, and race identifiers are some of the edits conducted to produce this Numident file. This work is conducted by the Census Bureau. The resulting Numident file contains the most recent name and DOB data. All alternate name and DOB data are kept on separate files.

[45] We use the most recent citizenship information for individuals on the file. SSA is not automatically notified when individuals naturalize though there are incentives to notify SSA after obtaining citizenship (Brown et al. 2018).

information is then used to determine whether the business owner is born in or outside the U.S..[46] In future work, whenever sex or age are not available in the Numident, previous census records data will be used whenever possible.[47]

Please note that while NES-D will follow ABS' questionnaire on place of birth and citizenship questions, the 2012 SBO questionnaire captured the concept of U.S. citizenship at birth, regardless of whether the respondent was born in the U.S. or abroad. Specifically the question asks "Was Owner # born a citizen of the United States?".[48]

Administrative Records Veterans Data

Administrative records from the Department of Veterans Affairs (VA) will supply the veteran status of the owner. Specifically, we use the U.S. Veterans Eligibility Trends and Statistics (USVETS) database, which is a comprehensive data source for veterans developed and maintained by the VA. The USVETS integrates data from multiple administrative records sources such as the Veterans Benefit Administration, the Veterans Health Administration, the Department of Defense's (DOD) Defense Manpower Data Center as well as other sources including commercial data sources. It contains every living and deceased U.S. veteran according to the VA's definition of a veteran as well as information such as demographics, details of military service, and VA benefit usage. In this project,

[46] Entrepreneurial activity of the individuals born outside the U.S. is an important and growing dimension of business dynamics. While entrepreneurial activity of those born outside the U.S. has been increasing, business start-ups have been declining in the U.S. and those start-ups are generating fewer jobs. For instance, individuals born outside the U.S. have higher business start-up rates than the native-born (Fairlie and Lofstrom, 2015), their businesses have grown steadily over the last decades and there is evidence that entrepreneurs born outside the U.S. have made important contributions in the technology and engineering sectors in the U.S. (Wadwha et al., 2007).
[47] Prior research on administrative records show high agreement rates between age and sex information in the Numident and previous census records. See, for instance, Bhaskar et al. (2014), and Rastogi & O'Hara (2012).
[48] The full 2012 long form questionnaire is available at: http://www2.census.gov/programssurveys/sbo/questionnaire/2012/2012-sbo-questionnaire/sbo1_2012.pdf.

veteran status assignment is based on the extract from the USVETS database for fiscal year 2015. We consider any nonemployer firm owner whose record is found in the USVETS database to be a veteran. We rely on the raw USVETS records without modifying or adjusting the data in any way.

Conceptual Differences of Veteran Status across Data Sources

Cross-data conceptual differences regarding what constitutes a veteran is an important issue that deserves special attention in the context of our work. In short, the veteran concept captured by the SBO (and the forthcoming ABS) is broader than the one identified by the VA's USVETS file. The VA defines a veteran as someone who has served on active military duty in the past. Accordingly, the VA does not consider individuals who are currently serving on active military duty as veterans, nor does it consider veterans those serving in the National Guard/Reserve component unless they also served on active military duty in the past.[49] By contrast, these two groups of individuals are considered veterans in the SBO and ABS surveys. Because of the VA's narrower definition of a veteran relative to the SBO and ABS, some individuals self-identified as veterans in the survey data are not part of the VA's AR data on veterans.

In particular, the 2012 SBO and the ABS surveys ask respondents (or their proxies) to self-identify their military service status.[50] Anyone who selects any of the categories below is classified as a veteran. Please note how the possible responses allow for never-activated reservists and also individuals who have served and continue serving at present – in contrast to the VA's definition of what constitutes a veteran. This broader identification of a veteran in the 2012 SBO, which

[49] Those who have never served on active military duty in the past are sometimes called "never-activated."
[50] See military service questions asked in the 2012 SBO and the ABS in Appendix 4.

subsequently carried over to the ABS, was originally based on feedback from multiple veteran group representatives. The categories are:

- ever served in any branch of the U.S. Armed Forces, including the Coast Guard, the National Guard, or Reserve component of any service branch
- disabled as the result of illness or injury incurred or aggravated during military service
- served on active duty during military service, not including training for the Reserves or National Guard
- served on active duty military service after September 11, 2001
- served on active duty military service in a given reference year
- served in the National Guard or as a reservist of any branch of the U.S. Armed Forces in a given reference year

Our research indicates that, despite some limitations, the VA's USVETS database is the most comprehensive administrative records source currently available that identifies veterans in the U.S. (based on the VA's definition of a veteran). However, because it only includes veterans according to the VA's definition, it limits our ability to identify veterans in a manner consistent with the surveys' broader veteran definition. We are currently exploring the possibility of employing an additional AR data source that may be able to complement the USVETS data. These data are the DOD's Defense Enrollment Eligibility Reporting System (DEERS) database. This would allow us to better align the SBO/ABS' definition of a veteran with the one we can obtain using AR data. Appendix 3 contains an extensive discussion of the DEERS data and its potential use in the assignment of veteran status. In addition, because the USVETS database is the primary data source we are considering to assign veteran status information to nonemployers and we need to identify first-hand any potential quality

issues with the data, we conduct some quality checks using other data sources of veteran status. We discuss our results from this analysis in detail in Appendix 5. In general, the USVETS data are less accurate for older and healthier veterans. Because of a 1973 fire at the National Personnel Records Center that destroyed approximately 16 to 18 million official military personnel files, older veterans are likely to be undercounted in the USVETS database.[51] As older veterans die, this limitation of the USVETS data will gradually decrease over time. Due to the USVETS reliance on various administrative records sources, veterans who have never used the VA medical facilities or who have never enrolled in the VA benefit programs are not captured by the USVETS data. As a result, healthier veterans may be undercounted in the current version of the USVETS database. As the USVETS database updated with data from additional administrative and commercial sources, this limitation should also decrease over time.

Potential/Future Data Sources

In future work we plan to expand the NES-D series with additional characteristics that are relevant to the understanding of nonemployers dynamics. Some of these characteristics would include household attributes, such as marital status, number of dependents or home ownership. Tax data from Form 1040 would be used to construct such characteristics. W-2 data can also provide valuable information on whether nonemployers' income is the primary source of income of nonemployer businesses. The Business Register, the Longitudinal Business Database (LBD) and the Integrated Longitudinal Business Database (ILBD) could be used to obtain the age of the business, and transitions from the nonemployer to the employer universe.[52] Other potential variables from

[51] See more information about this fire and its impact here: https://www.archives.gov/personnel-records-center/fire-1973.

[52] See https://www.census.gov/ces/dataproducts/economicdata.html for a description of the LBD and ILBD.

non-tax sources of high value for the NES-D are trade status from the Importer and Exporter Databases, and measures of patenting and innovation activity from the U.S. Patent and Trade Office (USPTO). The Exporter Database and Importer Database contain key data items from export/import statistics and firm identifiers from the BR. Export data include commodities exported, shipment value, weight, country of destination, U.S. customs port, and method of transportation. Import data include commodities imported, shipment value, weight, country of origin, U.S. customs port, and method of transportation.[53]

III. Methodology

In this section we review the methodology to identify nonemployer business owners by LFO, and discuss some methodological challenges in the assignment of demographics at the individual and firm level.

As mentioned above, we use individual anonymized identifiers or PIKs to obtain and attach demographic characteristics from a variety of sources to nonemployer business owners. The source of PIKs depends on the legal form of organization of the business. In the case of nonemployer sole proprietors, PIKs come from Form 1040 data in the BR. PIKs for partners and S-corp owners are obtained from Forms 1065 and 1120S Schedule K-1. K-1 data contain the firm's EIN along with the PIKs of the owners. Hence, for a given partnership or S-corp, we are able to obtain EIN-PIK pairs, which identify a given firm and all its owners.

The task of identifying owners of nonemployer C-corps is a lot more challenging, as there is no tax form or business registry that clearly identifies owners of C-corps in the U.S.. It is beyond the

[53] See https://www.census.gov/econ/overview/mt0300.html for more information.

scope of this initial report to identify these owners and characterize or impute demographics for this group of firms. Fortunately, as mentioned earlier, C-corps only represent 2 percent of the nonemployer population, and 4 percent of nonemployers receipts. It is worth mentioning that person-ownership of C-corps may be very diffuse and/or the C-corp itself may be owned by other companies. This will have implications for what C-corps are eligible for demographic characteristics imputation. We plan to explore all these issues in future work, which will include characterizing and taking a closer look at nonemployer C-corps by linking them to multiple years of the BR and prior SBOs to obtain information about their demographics, industry, geography and receipts. We hope this exploration will shed light into issues such as: are they new firms that are soon to be employers? Are they dying firms - that perhaps used to be employers? Are they mis-categorized employer firms? Do C-corps look like partnerships or S-corps in terms of their receipts, demographic, industry, geographic distributions?[54] This information will be useful to develop a demographics imputation model for these firms. If our research shows that imputation of demographics for C-corps cannot produce reliable estimates, we will provide transparency, and document and share our findings in a future paper.

Assigning demographic characteristics to owners of sole proprietorships, and by extension to the firms themselves, is straightforward. Only individuals can own sole proprietorships, and each

[54] It is also possible that some of these C-corps originated as sole proprietorships, partnerships or S-corps that later became C-corps for a variety of reasons. For these, we may be able to obtain past owner information via a business address match.

sole proprietorship has only one owner.[55] Any such firm owned by a person who can be matched to source data for a given demographic characteristic can be assigned that characteristic.

For partnerships and S-corps, the assignment of demographic characteristics is somewhat more complicated. Not all owners of these types of firms are necessarily individuals. Partnerships, for example, explicitly may be owned in part or in whole by other firms. While other firms may not as a general rule own S-corps, there are some circumstances in which this arrangement may be permitted, and we do observe it in some cases in the K-1 data.[56] Firms owned at least in part by other firms are called tiered entities.[57] A tiered partnership's Schedule K-1 will provide the identity of the other firms that own it, but the form will not provide the identities (i.e., PIKs) of the individual owners of those firms. Here, we do not attempt to identify higher-tier individual owners of nonemployer partnerships and S-corps. Instead, we focus on the demographic characteristics of the individual owners listed directly on the firms' K-1s.[58]

Just like in the SBO, NES-D will provide nonemployer estimates by firm-level demographics. While the assignment of demographic characteristics to identified individual owners of partnerships and S-corps is again straightforward, the aggregation of this information to the firm level is less so. Here we discuss several issues to consider when assigning demographics at firm

[55] As mentioned earlier, one exception to this one-owner rule is that married couples can legally own sole proprietorships jointly as "qualified joint ventures." Spouses who choose to have their businesses organized in this way must both participate in the business and file a joint income tax return, and the business may not have any members except for the couple.
[56] For example, an estate or trust may own shares of an S-corp for a limited time following the death of the individual who previously held them.
[57] Also see https://www.gpo.gov/fdsys/pkg/FR-2018-08-16/pdf/2018-17276.pdf.
[58] Approximately 20 percent of partnerships (about 2 percent of all nonemployer firms) in our 2015 sample are tiered entities. For half of those, people own 50 percent or more of the partnership. For more information on the characteristics of owners of all partnerships and S-corporations, including those with employees, see Goldschlag, Kim, and McCue (2017).

level and how to address them. Schedule K-1 includes ownership shares for each owner, which provides a reasonable way to allocate ownership of a firm to various demographic groups, but it provides these shares at two points in time – at the beginning and end of the year. If a firm changes owners during the year, it could, for example, be classified as male-owned at the beginning of January and female-owned at the end of December. How should such a firm be categorized for the year? In this work we use the average of the beginning- and end-of-year ownership shares for each owner to assign ownership to demographic categories.

There are also practical issues related to ownership shares that must be addressed before using them to aggregate demographic information.[59] First, the individual ownership shares reported by firms do not always account for 100 percent of the firm's ownership, and they sometimes account for more than 100 percent of the firm's ownership. This is uncommon, especially for S-corps, which have complete ownership represented by the owners listed on Schedule K-1 nearly 98 percent of the time; 89 percent of partnerships have their complete ownership listed on Schedule K-1. Notably, once owners that are themselves firms (and which do not figure into the demographic analysis) are excluded, it becomes more common for ownership shares to sum to less than 100 percent within firms. About three-quarters of partnerships that have at least one person as an owner have their full ownership reported on Schedule K-1 and attributable to owners who are people (rather than firms).

[59] A minor issue is that different firms fill out Schedule K-1 in different ways. Some of them use ownership shares that are between zero and one, while others use ownership shares between zero and 100.We address this primarily by rescaling ownerships shares within firms as discussed below. When identifying firms to exclude from categorization based on the "ten percent rule," we exclude those for which the largest owner's share is less than 0.1 or greater than one but less than ten.

Finally, there are questions about how many owners ought to figure into the demographic analysis and which firms ought to be categorized. When nonemployer statistics were produced from survey data in the SBO, practical considerations such as constraints on paper forms and a desire to limit respondent burden led to only the four largest owners being used to determine firm demographics. When the demographic information is drawn from administrative records, as it is here, neither of those constraints apply. How many owners should be used to determine firm demographics in the absence of these constraints? The survey-based version of these statistics also did not categorize firms in which the owner with the largest ownership share owned less than ten percent of the firm, again to reduce respondent burden. Moreover, it is not conceptually obvious that it makes sense to categorize firms in which ownership is very diffuse (e.g., where there is no owner with at least 10 percent ownership). Ownership shares have technical meaning for tax administration, but economically, they serve as proxies for control over the firm, which is not directly observable. Are firms with many small-share owners meaningfully controlled by any of them or any demographic group in particular? These questions do not have clear answers. Here, our main analysis follows the practices used in the legacy SBO. We use the four largest owners and do not categorize firms in which the largest individual ownership share is less than ten percent, but also present and discuss alternative approaches in Section VI that take fuller advantage of the fact that this project is based on administrative data and not constrained by respondent burden considerations.

Specifically, we assign firms to demographic groups by determining the total share of firm ownership held by individual members of each group. A firm is assigned to a given group if persons of that group collectively own a majority stake (more than 50 percent) in the firm. In order to

determine majority ownership, we rescale (or "rake") the ownership shares reported by each of the four largest owners by the sum of the ownership shares reported by these owners, and work with these rescaled shares, which sum to 100 percent by construction.[60] Again, following the SBO, only firms with at least one owner with an ownership share of 10 percent or higher are included.

To provide an illustrative example of how we assign firms to demographic groups, suppose a partnership is owned by four individuals, who each report owning 15 percent of the firm, and one other firm, which owns the remaining 40 percent. Three of the individuals are women, and one is a man. The largest individual ownership share exceeds ten percent, so we proceed to categorizing it. We first exclude the ownership share belonging to the other firm, focusing only on the four individual owners, who collectively own 60 percent of the firm. We rescale each person's 15 percent ownership share by the total ownership reported by the group, so each individual owns 15/60 = 25 percent of the person-owned portion of the firm. To assign this firm to a sex group, we add up the individual ownership shares belonging to women (75 percent) and the shares belonging to men, (25 percent). Because women represent majority ownership of this firm, the firm is classified as female-owned.

We repeat this process for other demographic characteristics. Those characteristics that have only two categories at the individual level (e.g., sex, Hispanic origin or veteran status) also have a third category at the firm level: equally-owned. For example, sex includes the following categories: female-owned, male-owned and equally-owned by men and women. For characteristics

[60] It is relatively rare for firms to have more than four owners. Among partnerships, 91.4 percent of firms report four or fewer person-owners on Schedule K-1, while 98.5 percent of S-corporations report four or fewer owners.

that have more than two individual-level categories, such as race, it is possible that no one group will collectively own a majority of each firm. Such firms are not assigned to groups for these demographic characteristics.[61] Missing demographic information at the individual level for some or all owners can also lead to a firm not being classified if that missingness prevents any group from reaching majority ownership.

Missing demographics will be imputed in future work in NES-D production. Demographics may be missing for a variety of reasons:

i) A PIK is not available for the owner. In the case of partnerships and S-corps, it is possible that the nonemployer firm cannot be found in K1 tax data, hence making it impossible to identify the owners of that business. It is also possible that K-1 data do not account for all owners or 100 percent ownership of the firm. In the case of sole proprietors, the PIK value may simply be missing,

ii) A PIK is available but some (or all) demographic characteristics are missing. This may happen whenever the PIK cannot be found in a given AR or census record source, or because the PIK matches but the value of the demographic characteristic is missing in the data source.

In Section V, we quantify demographic characteristics missingness in the nonemployer data universe and its sources by LFO.

[61] The SBO also grouped race categories by minority status and provided firm-level statistics by whether the firm was minority owned, non-minority owns or equally owned by minority and non-minority groups. We plan to add this classification in future work.

IV. PIK Assignment Results

Here we present findings on PIK assignment and coverage for owners of nonemployer sole proprietorships, partnerships and S-corps.[62] Regarding sole proprietors, we are able to identify owner PIKs for 99.9 percent of sole proprietorships (see Table 1).[63] Remember that for partnerships and S-corps, owner PIKs are identified from Schedule K-1, which reports all owners of each business, as well as the share of the business owned by each owner.[64] Using the business' EIN in the nonemployer database, we first match nonemployer firms operating in 2015 sequentially to data from Schedule K-1 2015, 2016, and 2014 to obtain EIN-PIK pairs (i.e., firm-owner pairs). As reported in Table 2, we are able to match the vast majority of partnerships (99.3 percent) and S-corps (99.1 percent) to K-1 data (with nearly 99 percent of our matches coming from 2015 K-1 data).[65] This means that we have PIK information for more than 99 percent of nonemployer businesses. However, there are some K-1 firms for which we cannot account for 100 percent of their ownership. These represent less than 1 percent of all nonemployer businesses. Owners reported on Schedule K-1 account for full ownership (i.e., their ownership shares sum to 100 percent) of 89 percent of partnerships and approximately 98 percent of S-corps. Firms for which the total

[62] Please note that because of rounding in compliance to current Census Bureau Disclosure Review Board's disclosure rules, in some cases columns or rows may not add up to the total. As already mentioned, all values in tables and figures have been rounded to four significant digits as part of the disclosure avoidance protocol. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000 and above 10,000,000 to the nearest 10,000.

[63] The remaining 0.1 percent of sole proprietorships do not have valid owner PIKs assigned in the file we use for this analysis. We are exploring ways to assign valid PIKs to these sole proprietorships.

[64] For partnerships, this is the partner's share of capital. For S-corporations, it is the shareholder's percentage of stock ownership. In general, Schedule K-1 is used to report the shares of a business's income, credits, deductions, etc. that belong to each of its owners. We assign ownership of firms according to these shares.

[65] This is when we allow for ownership shares to add up to between 99 and 101 percent (to account for possible rounding errors). When we require the sum of owner shares to add up exactly to 100 percent, the owners reported on Schedule K-1 account for full ownership of approximately 77 percent of partnerships and 96 percent of S-corporations.

reported ownership shares are not at or around 100 percent most commonly report total ownership around 50 percent (data not shown).[66]

Most partnerships and S-corps have a relatively small number of owners, as shown in Table 3, but a small share have many owners, with 3.3 percent of partnerships and 0.3 percent of S-corps having at least 10 owners. Even outside of such extreme cases, partnerships tend to have more owners than S-corps, and as a result, the number of owner-partnership pairs identified on K-1s is larger than the number of owner-S-corporation pairs. Partnerships have more owners whether owners that are themselves firms are included in the count or not.

From our match to K-1 data, we have identified about 6,687,000 firm-owner pairs in the partnership data, as reported in Table 4. Of those pairs, about 5,557,000 (83.1 percent) involve individual owners, while the other 1,130,000 (16.9 percent) involve firm owners (data not shown). There are about 1,632,000 owner-S-corp pairs, of which 1,603,000 (98.2 percent) involve individuals and 29,000 (1.8 percent) other firms (data not shown). As this suggests, partnerships are much more likely than S-corps to be tiered entities. This is to be expected since, as explained earlier, other firms may not as a general rule own S-corps – except in special circumstances. As shown in Table 5 column 9, among partnerships, 23.5 percent have at least one owner that is itself a firm, while only 1.7 percent of S-corps are tiered entities. Tiered entities though only represent about 2 percent of the nonemployer universe, and for about half of them, persons own 50 percent or more of the business. In addition, for about 40 percent of tiered entities, the largest ownership share owned by an individual is less than 10 percent. Note that following SBO's methodology, those

[66] We plan to quantify and document this issue more thoroughly in upcoming work.

firms are not eligible to be assigned demographics - and moreover it is not clear if it makes sense to demographically categorize firms with very diffuse ownership.

Table 6 reports the number of unique owners of nonemployer firms, organized by the type(s) of firm owned. Note that it is possible for individuals to own multiple nonemployer firms, including multiple nonemployer firms that have different legal forms of organization. The vast majority of nonemployer firm owners own only one type of firm, with sole proprietorships being by far the most prevalent, followed by partnerships and then S-corps. Individuals who own multiple types of firms are most likely to own at least one sole proprietorship and at least one partnership, with that arrangement being more than twice as common as owning sole proprietorships and S-corps. Only 0.6 percent of people who own nonemployer firms own partnerships and S-corps. Very few people own all three types of firms.

V. Availability of Demographics & Prevalence of Missingness Results

Here we discuss AR availability and coverage of demographic characteristics for the universe of nonemployers by LFO, and quantify the prevalence of missing values across demographic characteristics.

A. Match to Census Numident: Sex, Age, Place of Birth and Citizenship

Sex and Age

Table 7 reports match rates to the 2015 Census Numident and availability of sex and age information for firm-owner pairs (where owners of multiple firms are counted multiple times) by legal form of organization. Match rates to the Numident are very high – over 99 percent of sole proprietorship and S-corps firm-owner pairs, and 97 percent of partnership firm-owner pairs match

to the Numident. Sex and age information is essentially universally available for individuals who are matched to the Numident.

Place of Birth and Citizenship

Table 8 shows availability of place of birth and citizenship information for firm-owner pairs by legal form of organization. Place of birth information in the Numident is based on two variables – one which indicates a two letter code for the state or country of birth and a second which indicates whether that two letter code refers to a) a U.S. state, Puerto Rico, or a U.S. island area, or b) another country. Availability of place of birth data is shown in Columns 3a and 3b of Table 8. For a small number of cases, the state/country of birth code in the Numident is missing so we cannot determine if these individuals were born in or outside the United States; however, overall, place of birth data are available for the majority of firm-owner pairs. Specifically, the Numident provides place of birth information for 99.4 percent of sole proprietorship firm-owner pairs, 97.4 percent of partnership firm-owner pairs, and 99.4 percent of S-corps firm-owner pairs.

Citizenship information is based on a variable in the Numident which is missing for about 25 percent of firm-owner pairs. However, an evaluation of the missing responses suggest we can infer citizenship for many of these cases. To evaluate these missing responses, we linked the data to previous census records which contained reported citizenship information from the Census 2000 long form and American Community Survey for years 2001 through 2011. We found that among firm-owner pairs with missing citizenship information in the Census Numident and for whom we had previous census responses, almost all (99.3 percent) had been reported as being U.S. citizens in the previous census responses. Therefore, one option for dealing with cases that match to the Numident but have missing citizenship data is to assume that all these individuals are U.S. citizens.

This is shown as Option A in Table 8 (columns 4a and 4b). Using this assumption, all firm-owner pairs that match to the Numident are assigned citizenship status, resulting in availability for between 97 to 99 percent of firm-owner pairs depending on LFO. Recent work by Brown et al. (2018) further evaluated citizenship data in the Numident and develop an alternative method of dealing with missing cases. For most cases with missing citizenship data, they also assume the individuals are U.S. citizens, but they consider those with missing citizenship whose place of birth data in the Numident indicates a non-U.S. place of birth to have missing citizenship information. We show the results of using this assumption as Option B in Table 8, and as expected we find slightly lower rates of availability of citizenship compared to Option A. Using Option B, about 98 percent of firm-owner pairs have citizenship data. Citizenship is available for 98.4 percent of sole proprietorship firm-owner pairs, 94.9 percent of partnership firm-owner pairs, and 96.8 percent of S-corps firm-owner pairs. Since the two options produce similar results, and based on our own as well as Brown et al. (2018) research, we focus our discussion and results on Option B, where we consider those with missing citizen values in the Numident and a place of birth outside the United States to have missing citizenship information. By contrast, those with missing citizenship values but a place of birth in the United States are considered to be U.S. citizens.

B. Match to Previous Census Records: Race and Ethnicity

As mentioned earlier, race and ethnicity results presented here focus on the match of 2015 PCR file with the nonemployer data. We are also evaluating the potential use of an Administrative Records Composite file for assigning race and ethnicity, and we have an extensive discussion of those results in Appendix 2.

Table 9 provides match rates for the nonemployer firm-owner pair universe linked to the 2015 PCR file. Match rates and availability of race and ethnicity demographic information are shown by legal form of organization. Overall, there is a high match rate (approximately 94 percent) between firm-owner pairs and PCR data. More specifically, the match rate for sole proprietorships is about 93 percent while for partnerships and S-corps is about 95 and 96 percent respectively.

In order to adhere to the revisions to Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting issued by the Office of Management and Budget, NES-D will not allow for a race category of "Some Other Race".[67] Here we show in just Table 9 the availability of Some Other Race responses as these will eventually be reclassified. The rest of the race tables and analyses in the paper will not include the Some Other Race category. In future work, we plan to employ a Census algorithm to redistribute responses of Some Other Race to the other allowed race categories. [68]

Table 9 Column 3a shows the availability of race responses in the PCR data file as a percentage of the firm-owner pairs. Race responses are available for about 88.5 percent of sole proprietorships, 94.1 percent of partnerships, and 94.0 percent of S-corp firm-owner pairs. Note that these percentages are for all firm-owner pairs. Firm-owner pairs that cannot be linked to the PCR file will be assigned an imputed race response in future work. If we consider only the firm-owner pairs linked to the PCR file, about 95 percent of the firm-owner pairs matched to this file

[67] The Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity were issued in 1997 and can be accessed at https://www.whitehouse.gov/wp-content/uploads/2017/11/Revisions-to-the-Standards-for-the-Classification-of-Federal-Data-on-Race-and-Ethnicity-October30-1997.pdf.

[68] According to the "Overview of Race and Hispanic Origin: 2010," Hispanics made up 97 percent of all those classified as Some Other Race alone in the 2010 Census.
https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf

have a race response available, and between 1 and 4 percent have a race response of Some Other Race.

For ethnicity, Hispanic origin responses are available for all firm-owner pairs matched to the PCR file (column 2a and column 5a in Table 9). Column 5b shows that 93.0 percent of sole proprietorships, 94.9 percent of partnerships, and 95.8 percent of S-corps can be assigned Hispanic origin using PCR data. Hispanic origin information will be imputed for firm-owner pairs that either cannot be found in the PCR file or can be found but have missing Hispanic origin information.

C. Match to VA Administrative Records Data (USVETS)

As with our other demographics, we assign veteran status by linking nonemployers to the USVETS database via PIK. However, in the case of our match to the USVETS, there is no meaningful way of assessing and assigning missingness status. This is because the USVETS contains only individuals who have been identified as veterans by the VA; therefore, a match to the USVETS file simply implies that that particular business owner is considered a veteran by the VA while a non-match implies that the person is not considered a veteran.

D. Prevalence of Missingness

Since individual-level demographic missingness has implications for our ability to classify owners and firms by demographic group, here we discuss the prevalence of missing demographic characteristics among our nonemployers. As mentioned earlier, demographics may be missing for a variety of reasons: inability to assign a PIK to a particular owner, the PIK cannot be found in a given demographics AR or census data source, or the value for a given demographic characteristic may be missing in the AR or census data.

Firms with EINs that cannot be matched to K-1 data have no PIK or demographic information available for their owners. As a result, they will need to have both the number of owners and the characteristics of those owners imputed. Fortunately, as shown in Table 2, only approximately 23,000 K-1 firms (less than 1 percent) are in this situation. Also luckily, only 20,000 nonemployer sole proprietors (0.1 percent) have a non-valid PIK, and we are exploring ways to obtain a valid PIK for those cases.

Looking now at nonemployers with PIKs, as shown in Table 10, the vast majority (over 90 percent) of nonemployers are not missing any demographic characteristic, while only about one percent are missing three or more. Because both race and Hispanic origin are obtained from PCR data, when either one is missing, the other is generally also missing (see Table 11). This sometimes occurs in combination with at least one missing characteristic from the Numident.

At the individual level, race is the most commonly missing piece of demographic information, followed closely by Hispanic origin; each is missing for about seven percent of owners, as shown in Table 12. This is because information for both variables originates in the PCR file. Citizenship is missing for about 2.3 percent of nonemployer firm owners, making it the demographic characteristic from the Numident with most missing values. The Social Security Administration began requiring documentation of citizenship status as part of the SSN application process in 1972 (Brown et al. 2018). For this research we developed assumptions, as described above, which allow us to assign citizenship status to most of the firm owners with a blank citizenship code in the Numident. Other demographic information obtained from the Numident is missing for only about 1 percent of owners. Veteran status data are not missing for any owners

because, by design, absence from the USVETS data indicates that one is not a veteran – although there coverage issues as previously discussed.[69]

VI. Business Owner and Firm-level Demographics Results

This section presents our results on nonemployer demographics at firm-owner, unique owner and firm levels for all industries and at the national level. Remember that a given individual can own more than one business, so the firm-owner tables take into account this fact and, by construction, give more weight to demographics of individuals that own more than one business.[70] The unique-owner results represent a profile of the U.S. entrepreneur, regardless of how many businesses he/she owns. Firm-level tabulations follow the methodology outlined in the Methodology section and give us a sense of firm-level diversity and demographics. These initial tabulations should be interpreted with caution since they do not include imputed values of missing demographics, and are incomplete. Therefore, at this early stage, they are not intended to be representative of the demographics of the underlying nonemployer population. Although these preliminary results are not fully comparable to prior SBO publications, we nevertheless undertake an initial comparison to see if they behave according to our expectations, and overall they do – as we discuss below.

[69] Missingness also has implications for assigning demographics at the firm level. As shown in Table 13, it is fairly uncommon, however, for all K-1 firm owners to be missing demographic information. Only about 1.2 percent of partnerships have all owners missing some demographic data, and only 0.6 percent have all owners missing all demographic data. For S-corporations, about 0.4 percent are missing some demographics for all owners, while about 0.3 percent are missing all demographics for all owners. As stated earlier, though, we plan to impute missing demographic characteristics in upcoming work.

[70] In firm-owner counts, a given demographic is given a weight of 1 regardless of the owner's ownership share in the firm.

A. Sex

At the firm-owner level, men represent the majority of owners of nonemployer firms for all LFOs, as shown in Table 14. About 55 percent of sole proprietors are men, while about 45 percent are women. Among both partnership and S-corporation owners, about two-thirds are men and one-third are women.

Since individuals can own multiple firms, the previous estimate essentially weights people according to the number of nonemployer firms that they have some ownership of (though not by the degree of ownership in those firms). When each owner is counted only once, no matter how many or which type(s) of firms he or she owns, about 56 percent of nonemployer owners are men, and 43 percent are women, as shown in Table 15. These figures are very similar to the firm-owner estimates since the vast majority of nonemployers are sole proprietors and the vast majority of nonemployer owners own exactly one firm. Table 16 shows that about 90 percent of nonemployer firm owners own exactly one such business; another 7.3 percent own two businesses, and 1.3 percent own three, leaving less than one percent of owners holding more than three nonemployer businesses. Information on owners sex could not be matched to the remaining one percent of owners.

As discussed in the Methodology section, our primary analysis follows the conventions of the legacy SBO with regard to how firms are assigned to demographic categories. Remember that in the case of firms with multiple owners, the SBO used only up to the four person-owners with the largest ownership shares in each firm, and did not assign demographics to firms in which the person-owner with the largest ownership share owns less than ten percent of the firm. For the

SBO, these conventions were based in large part on practical considerations such as limiting respondent burden and preserving space on survey forms. However, when assigning firms to demographic categories using administrative records that have already been collected, these considerations do not necessarily apply. We could, in theory, use administrative data on all person owners of every firm to assign demographic information at the firm level. This possibility raises the question of whether using information on more owners or categorizing more firms would meaningfully change the shares of firms falling into various demographic categories. The scope for these changes to affect firm categorization is substantially limited by the fact that the vast majority of firms in question have four or fewer owners.[71] However, the degree to which firms with more owners might be affected is an empirical question. An additional question (as discussed in the Methodology section) is whether it makes sense to categorize firms in which the person-owner with the largest share owns less than 10 percent of the firm (i.e., businesses where ownership is very diffuse). Are firms with many small-share owners meaningfully controlled by any of them or any demographic group in particular?

Table 17 reports rates of firm ownership by sex categories under our baseline approach to categorization as well as three alternative approaches: 1) using the four largest owners and not imposing the ten percent rule, 2) using all owners and imposing the ten percent rule, and 3) using all owners and not imposing the ten percent rule. Results are presented in aggregate and by legal form of organization.

[71] Refer to Table 3 to see the distribution of number of owners for partnerships and S-corps.

We see (in Table 17) that sole proprietorships are simply assigned the sex of their owners, so about 55 percent of sole proprietorships are male-owned, and nearly 45 percent are female-owned. Following first the baseline/SBO approach, partnerships and S-corps are placed into one of three categories based on majority ownership: male-owned, female-owned, and equally male and female-owned (the male and female ownerships shares each equal 50 percent). Among partnerships, about 52 percent are classified as male-owned, 15 percent as female-owned, and 26 percent as equally owned. S-corps are also most likely to be male-owned, with about 64 percent falling into that category. Just over 23 percent of S-corps are female-owned, and 12 percent are equally owned. For both types of firms, the remainder could not be classified under the scheme we have selected for the main analysis here, either due to the imposition of the ten percent rule or due to missing data on owners.

The first-order implications of the alternative approaches we consider are most directly observable in the "Not Assigned" columns. For partnerships and S-corps combined, dropping the ten percent rule would lead to the categorizations of an additional 2.5 percent of firms. This change is driven almost entirely by partnerships, where dropping the ten percent rule would lead to the categorization by sex of an additional 4 percent of firms; there is very little change in assignment rates for S-corps. The implications of using all owners instead of the four largest are substantially smaller than those of relaxing the ten percent rule. Moving from our baseline approach to the approach that uses all owners and imposes the ten percent rule leads to successful sex assignment for an additional 0.7 percent of firms.

Table 17 also shows that the changes in the distribution of firms across sex categories associated with these alternative approaches are generally small, especially for S-corps. Within

partnerships, the number of male-owned firms increases by more than the other categories when we drop the ten percent rule from the four-owner approach, for example, suggesting that firms on this margin tend to be male-owned. Partnerships in general tend to be male-owned (about 52 percent under our baseline approach). Therefore, the firms on this assignment margin are both mostly and disproportionately male-owned. However, the implications of this for the overall firm sex distribution are small because the share of firms on this assignment margin is small. [72]

Comparison to SBO

Aggregating across legal form of organization, we can contrast our firm-level administrative records-based sex classifications from the baseline approach to survey-based measures from the 2012 SBO. As already mentioned, our preliminary results should be interpreted with caution and are not intended to be representative of the demographics of the underlying nonemployer population. We nevertheless make an initial comparison to prior SBOs to see if our AR-based estimates behave according to expectations.

As shown in Table 18, the female and male-owned estimates are broadly similar. The clearest difference is in the share of firms that are equally owned by men and women, with our AR-based estimate being notably lower. This is to a great extent due to the fact that we classify all sole proprietorships as either male-owned or female-owned, since we use the sex of the single owner identified on their tax forms. Since the vast majority of nonemployer businesses (86 percent in 2015) are sole proprietorships, the AR-based percentage of nonemployer businesses equally owned

[72] Some preliminary exploratory work using revenue-weighted counts indicate no significant deviations from the ones presented here.

by men and women is lower than the percentage obtained from the SBO.[73] Note though that changes of similar magnitudes in the share of firms equally owned by men and women occurred between previous rounds of the SBO. This is partly attributable to the fact that the 2012 SBO used administrative records data to direct replace or impute (whenever direct replacement was not possible) sex, race, Hispanic origin and veteran status for nonemployer sole proprietors.

B. Race and Hispanic Origin

Table 19 shows the distribution of race responses for all firm-owner pairs after the match to the 2015 PCR data. In this table, both firm-owner pairs with a match and those without a match are included – while Table 20 presents results that exclude missing race values. From Table 19 we see that about 10.1 percent of all firm-owner pairs are missing a race response, 74.2 percent fall under the White alone category, 8.4 percent under Black alone, and 5.4 percent under Asian alone. Multiple races account for 1.6 percent of the distribution, while American Indian or Alaska Native (AIAN) and Native Hawaiian or Other Pacific Islander (NHPI) account for less than 0.2 percent.

We also see some variation across legal form of organization. Sole proprietorships have a lower share of White alone (70.7 percent) than partnerships (85.3 percent) and S-corps (81.2 percent), and a higher share of Black alone responses (10.6 percent) than partnerships (1.5 percent) and S-corps (3.6 percent). The share for other groups is similar across legal form of organization.

[73] As mentioned in the Methodology section, married couples can legally jointly own a sole proprietorship if they file taxes as a "qualified joint venture". Future work will explore the identification of these cases through tax data.

In NES-D, any firm-owner pairs with a missing race will have one assigned through imputation or the use of an algorithm. Since our estimates do not yet include imputation and the percent of missing race values is not trivial (representing about 10 percent of firm-owner pairs), in Table 20 we also present race distribution results when missing race values are excluded. Overall, we see that 82.5 percent of firm-owner pairs matched to PCR data now have a race response of White alone, 9.3 percent have a Black alone response, and 6.1 percent have a race response of Asian. Multiple races now account for 1.8 percent of the available race responses, and AIAN and NHPI account for 0.2 percent or less. We still see variation across legal form of organization when missing race values are excluded (Table 20). Among sole proprietorships, 79.9 percent are White alone in the PCR data file and 12.0 percent are Black alone. The share of White alone is higher and the share of Black alone is lower for partnerships and S-corp firm-owner pairs. Among partnerships, 90.6 percent are assigned White alone and 1.6 percent are assigned Black alone. Among S-corps, 87.1 percent of firm-owner pairs are assigned White alone and 3.8 percent are assigned Black alone. The share of Asian-owned businesses is similar across legal forms of organizations: sole proprietorships (5.8 percent), partnerships (6.6 percent), and S-corps (7.6 percent). Multiple races, AIAN, and NHPI combined make up about 2 percent or fewer of the firm-owner pairs across legal form of organization.

Moving on to Hispanic origin, Table 21 presents firm-owner results that include cases with missing Hispanic origin values while Table 22 excludes them. We see that about 6.5 percent of all firm-owner pairs did not match to the PCR file and have a missing Hispanic origin response. Overall, 83.1 percent of the firm-owner pairs are assigned a non-Hispanic response, and 10.4 percent are assigned as being Hispanic. The share of non-Hispanics is lowest for sole proprietorships (80.6

percent) and highest for partnerships (91.4 percent), while Hispanics have the highest share for sole proprietorships (12.4 percent) and the lowest for partnerships (3.5 percent).

As with race responses, to see how the Hispanic origin distribution looks when missing values are removed, Table 22 presents the distribution of Hispanic origin responses for firm-owner pairs matched to the PCR file. Overall, after removing the missing cases, 88.9 percent of nonemployer firm-owner pairs are assigned to be non-Hispanic, and 11.1 percent as Hispanic. As with race responses, there is still variation across the legal form of organization. Partnerships have the highest share of non-Hispanics (96.3 percent) followed by S-corps (91.5 percent) and sole proprietorships (89.7 percent).

Table 23 shows match rates and availability of race and Hispanic origin responses for unique business owners according to legal form of organization. Column 1 shows the number of owners with unique PIKs for each type and combination of legal form of organization, and Column 5 shows the percent of unique firm-owner pairs that can be assigned a race response from the PCR file. Unique PIK owners of partnerships and S-corps have the highest availability of race data (97.5 percent), and unique owners of sole proprietorships only have the lowest availability (88.1 percent). These groups also have the highest and lowest match rate to the PCR file. For Hispanic origin, all records that match to the PCR file can be assigned Hispanic origin. Owners of partnerships and S-corps again have the highest availability of Hispanic origin responses (98.1 percent), and owners of sole proprietorships only have the lowest availability (92.8 percent).

The race distribution for unique business owners according to legal form of organization is shown in Table 24. We see again variation in the distribution of race responses according to legal

form of organization. As with the distribution for all firm-owner pairs, sole proprietorships have the lowest share of White alone (69.9 percent) and the highest share of Black alone responses (11.0 percent). The highest share of White alone (88.1 percent) and the lowest share of Black alone (1.1 percent) are seen among owners of partnerships and S-corps. The other race responses are more similar across legal form of organization owned.

Table 25 presents the Hispanic origin distribution for unique business owners by legal form of organization (with missing Hispanic origin values included). As with the distribution for all firm-owner pairs, sole proprietorships have the highest share of unique business owners with a Hispanic response (12.7 percent) and the lowest share of non-Hispanic responses (80.1 percent). Unique owners of partnerships and S-corps have the lowest share of Hispanic responses (3.8 percent) and the highest share of non-Hispanic responses (94.1 percent).

In Table 26 we turn to results for the assignment of race responses at the firm level. Results are shown using the previously discussed rules for assigning race to firms: using four owners with and without the ten percent rule, and using all owners with and without the ten percent rule. Recall that a firm is assigned to a given race or ethnicity group if persons of that group collectively own a majority stake (more than 50 percent) in the firm.

As with sex, some firms cannot be assigned a race response because no one race group collectively owns a majority of the firm. In addition, inability to link a given PIK to PCR data or missing race information (in the PCR file) at the individual level for some or all owners can also lead to a firm not being classified if that missingness prevents any group from reaching majority ownership. Again, the first-order implications of the alternative assignment approaches are most

directly observable in the "Not Assigned" columns. Overall, when considering all nonemployer firms, the method of assignment virtually has no effect on the percent of firms that are assigned a race. This is because the vast majority of nonemployer firms only have one owner. For partnerships though, dropping the ten percent rule would lead to the race categorization of an additional 3 percent of firms. There is virtually no change in assignment rates for S-corps. Also, just as with sex, the implications for race assignment of using all owners instead of the four largest are virtually non-existent. Specifically, using the 10 percent rule, 9.9 percent of partnerships are not assigned a race when using four owners (column 15), and 9.6 percent are not assigned a race when using all owners. This figure declines to 6.9 percent when the 10 percent rule is not applied using four owners and 6.8 percent using ten owners as more firms have one race group with a majority stake.

In regards to firm race distribution, overall, using either four owners or all owners, with or without the 10 percent rule being applied, about 72 percent of all firms are White-owned, 9.6 percent are Black-owned, 5.2 percent are Asian-owned, and 1.7 percent are designated as owned by multiple races. Fewer than two-tenths of a percent of firms are AIAN- or NHPI-owned. About 5 percent of firms are not assigned a race (column 15, Table 26), and another 6.5 percent of firms do not have any owners that match to the PCR data file (column 17, Table 26).

Continuing with Table 26, if we look at the firm ownership race distribution according to legal form of organization, we see that fewer sole proprietorships are White-owned (70.7 percent) relative to partnerships (81.1 – 84.0 percent) and S-corps (79.7 – 79.8 percent). Relative to partnerships and S-corps, more sole proprietorships are Black-owned, AIAN-owned, NHPI-owned and multiple races-owned. Asian firms are similarly prevalent across legal form of organization.

Table 27 shows the firm ownership Hispanic origin distribution (including missings). When it comes to differences due to the method of assignment, we observe the same pattern as in race. That is, the assignment approach has some impact on partnerships, but not on other legal forms. Specifically, for partnerships, we see that about three percent more partnerships have a Hispanic origin response assigned when the 10 percent rule is not used compared to when the 10 percent rule is applied. Overall, fewer than one percent of firms cannot be assigned a Hispanic origin response, and 6.5 percent of firms do not have any owners that match to the PCR data file.

In regards to firm Hispanic-origin distribution, more sole proprietorships are Hispanic-owned (12.4 percent) relative to partnerships (3.1 - 3.2 percent) and S-corps (8.6 – 8.7 percent), whereas fewer sole proprietorships are non-Hispanic-owned (80.6 percent) relative to partnerships (87.0 – 90.0 percent) and S-corps (85.3 – 85.4 percent). Overall, using either set of rules to assign Hispanic origin, 11.5 percent of firms are Hispanic-owned, about 81 percent are non-Hispanic-owned, and fewer than one percent are assigned equal ownership.

Comparison to SBO

By aggregating race and Hispanic origin across legal form of organization, we can contrast the race distribution of firms as assigned by the PCR file with survey-based measures from prior SBOs (see Table 28 for race and Table 29 for Hispanic origin). As already mentioned, a few precautionary notes are in order when looking at these results: i) Our estimates are calculated excluding the approximately 10 percent of owners with missing race or Hispanic origin information, ii) some of the observed differences will be due to time trend changes, iii) the 2007 and 2012 SBOs differed in the way they categorize the race of individuals that entered a Hispanic or Latino

response in the race write-in boxes,[74] iii) the SBO included a "Some-Other-Race" category while we do not,[75] iii) we include a "Multiple-Race" category while the 2012 SBO did not.[76] In the SBO, individuals of multiple races were assigned into their corresponding race categories. For instance, an owner who reported to be both Asian and White was counted separately as Asian and White in the SBO tabulations. As a result, in the SBO, businesses could be tabulated in more than one racial group.[77] In future work, we will include imputation of missing values and plan to adhere to SBO and ABS methodology to assign owners and firms in the multiple race category to other race categories. The current tables provide a sense of the frequency that the algorithm will need to be used to assign multiple races to other race categories.

As shown in Table 28, using the 2015 PCR data to assign race to firms, there are fewer firms for every race category. This is a consequence of approximately 11 percent of the 2015 firms not being assigned a race. Therefore, Table 30 shows the distribution using only firms that were able to be assigned a race category. Keeping in mind the cautionary notes from above, we see that the shares of White-owned, Black-owned, and Asian-owned firms from our 2015 AR-based estimation are within the values of the 2002/2007 SBOs and the 2012 SBO. Specifically, using the 2015 PCR data file, 81.1 percent are White-owned, which is more than the 2012 SBO (76.3 percent) but less than the 2002 (87.9 percent) and 2007 SBO (84.6). In the 2015 data, there are fewer Black-owned

[74] In the 2012 SBO, if a respondent entered a Hispanic or Latino ethnicity in the race write-in box, the record was categorized as "Some-Other-Race". By contrast, in the 2007 SBO, that same case would have been categorized as "White". The change was implemented to be consistent with 2010 Census methodology.
[75] In order to adhere to the revisions to Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting issued by the Office of Management and Budget, NES-D will not allow for a race category of "Some Other Race".
[76] Approximately 2 percent of owners were of multiple race.
[77] This could be because a sole owner reported to be of more than one race, a majority owner reported to be of more than one race, or because a majority combination of owners was reported to be of more than one race. See https://www.census.gov/programs-surveys/sbo/technical-documentation/methodology/2012-sbo-methodology.html .

firms (10.9 percent) than in the 2012 SBO (11.1 percent) but more than the 2002 (6.4 percent) and 2007 SBO (8.6 percent). The share of Asian-owned firms is also lower in 2015 (5.9 percent) than in 2012 (6.4 percent) but higher than in 2002 (4.5 percent) and 2007 (5.4 percent). Fewer firms are designated as AIAN-owned and NHPI-owned using the PCR data file compared to the SBO, and 1.9 percent of the 2015 firms are assigned as multiple race-owned.

Table 29 shows the same type of comparison for Hispanic origin, but this time we include firms that we have been unable to assign to a Hispanic origin category. The share of Hispanic-owned firms is comparable to the share in the SBO; however, because we include unassigned firms in this table, the share of non-Hispanic-owned firms is lower compared to the SBO. In Table 31, we show only the firms that were able to be assigned a Hispanic origin category. In this table, using either the 10 percent rule or not, we see that the share of Hispanic-owned firms using PCR data file (12.4 percent) is similar to the 2012 SBO (13.8 percent). The share of non-Hispanic-owned firms (87.6 percent) is similar to the share in the 2012 SBO (86.3 percent).

C. Veteran Status

Table 32 reports firm-owner level veteran status information by firm LFO. Individuals matched to the VA USVETS data are considered veterans, while those who did not match the VA USVETS data are considered non-veterans. As indicated above, in firm-owner level tables, individuals appear once for each firm they own, so that individuals who own multiple types of firms appear in multiple categories. Among the nonemployer firm-owner types we consider, we find that 6.3 percent of them are veterans based on the match to the USVETS data. This rate varies slightly

depending on the nonemployer LFO ranging from 6.2 percent among the sole proprietor firm-owner pairs to 6.9 percent among the S-corporation firm-owner pairs.

Table 33 reports owner-level veteran information by LFO of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. As before, we consider individuals matched to the VA USVETS data to be veterans and those who do not match the VA USVETS data to be non-veterans. We find that approximately 6.3 percent of nonemployer owners are veterans but the rate varies by LFO. Individuals owning firms of every LFO type and owners of both sole proprietorships and S-corps have the lowest veteran rate of 6 percent, while owners of only S-corps have the highest veteran rate of 7.6 percent.

Table 34 reports firm-level ownerships by veteran status and method of assignment for partnerships, S-corps, and sole proprietorships. Using the methodology explained in Section III, for partnerships and S-corps, a given firm is assigned to a veteran status group if members of that group account for more than 50 percent of ownership reported by either the four largest owners or by all owners, depending on the assignment method. Firms in which ownership is split 50-50 between veterans and non-veterans are assigned to the "Equal" group. Firms in which veterans do not account for more than 50 percent of ownership are assigned to non-veteran group.

Just like with the other demographics, results are shown using the previously discussed rules for assigning race to firms: using four owners with and without the ten percent rule, and using all owners with and without the ten percent rule. We find that at the firm-level partnerships have the lowest veteran rate of approximately 2.7 percent followed by S-corps with approximately 5.8 percent. Sole proprietorships have the highest veteran rate at 6.2 percent. We also find that

approximately 5 percent of partnerships are classified in the equal veteran-non-veteran category, while only slightly more than 2 percent of S-corps are classified in this equal category. Finally, veteran status cannot be assigned for more than 4 percent of partnerships and for 0.4 percent of S-corps (resulting in approximately 2.8 percent of partnerships and S-corps not being assigned). Just with other demographics, moving from four owners to all owners, both with the ten percent rule, results in almost no change in the veteran status assignment at the firm-level. The most notable difference in the firm-level assignment of veteran status is that removing the 10 percent rule, regardless of the number of owners used, reallocates firms from the missing category to the non-veteran category.

Comparison to SBO

As explained earlier, we expect our VA AR-based veteran estimates to be lower than the SBO's since the VA has a narrower definition of a veteran relative to the SBO. Table 35 contrasts firm-level ownership by veteran status based on data obtained from 2007 and 2012 SBO published tables (columns (1) through (4)) and 2015 VA USVETS data (columns (5) through (12)). The VA USVETS data columns include owners of nonemployer sole proprietorships, partnerships, and S-corps. We find that approximately 5.9 percent of all nonemployer firm types we consider are classified as veterans, regardless of the assignment measure we use. As expected, this rate is lower relative to the veteran rate from the previous SBO surveys because of VA's narrower definition of a veteran. This means that the USVETS, our veteran status source data, does not view as veterans some military personnel ; namely, those currently on active duty and those serving in the National Guard/Reserve component who never served on active duty in the past. As mentioned above, we will continue to explore additional sources of veteran data, such as administrative data from the

DOD, to fill the gaps in the USVETS data and to bring the identification of veterans closer to the SBO and ABS definition.

We find that approximately 0.4 percent of nonemployer firms are classified as equal veteran-non-veterans and this rate is much lower relative to the previous SBO data. We cannot assign a veteran status for about 0.3 percent of nonemployer firms because of the 10 percent rule. Removing the 10 percent rule does not cause significant changes in the veteran status assignment across all nonemployer firms.

D. Place of Birth and Citizenship

Table 36 and Table 37 show the place of birth and citizenship distribution for firm-owner pairs for whom data are available (using Option B as stated above for citizenship[78]) as well as the percent with missing place of birth and citizenship. A small portion of firm-owner pairs are missing place of birth (1.0 percent) or citizenship (2.4 percent); this indicates the individuals either did not match to the Numident or did match but place of birth or citizenship status was not available.[79] Overall, about 20.2 percent of nonemployer firm-owner pairs were born outside the United States (Table 36) and about 12.5 percent were not U.S. citizens (Table 37). This varies by legal form of organization. For sole-proprietorships, we find that 21.6 percent of firm-owner pairs were born outside the United States and 14.2 percent were non-U.S. citizens. For partnerships, we find lower rates of firm-owner pairs born outside the U.S. – about 13.9 percent of partnership firm-owner

[78] For reference, we show results of this table using Option A in Appendix Table A. 17. The results are very similar to what is shown in Table 37.

[79] As mentioned earlier, individuals with missing citizenship data are assumed to be U.S. citizens unless Numident information indicates that they are born outside the United States, in which case we classify them as having missing citizenship. Also see Brown et al. (2018).

pairs were born outside the U.S. and 6.4 percent were not U.S. citizens. Among S-corps, we find that 23.0 percent of firm-owner pairs were born outside the U.S. and 12.2 percent were not U.S. citizens.

Table 38 and Table 39 show the place of birth and citizenship distribution when each owner is counted only once, no matter how many or which type(s) of firms he or she owns. Once again a small portion of individuals are missing place of birth or citizenship data. Overall, about 20.5 percent were born outside the United States and 13.0 percent of owners are non-U.S. citizen. Most owners own only sole-proprietorships and thus we see similar patterns of place of birth and citizenship status for this group. As with the firm-owner tables described above, we find that a lower percentage of owners of partnerships were not U.S. citizens or born outside the United States. Nonemployer business owners who owned both sole-proprietorships and S-corps (a relatively small portion of our universe) were more likely than other types of owners to be born outside the United States (31.3 percent) or be non-U.S. citizens.

Next we turn to firm-level place of birth and citizenship status for partnerships and S-corps using the same assignment methodologies outlined earlier in the Methodology section. As shown in Table 40, depending on which methodology is used to assign firm-level place of birth between 17.0 and 17.3 percent of the 2.8 million partnerships and S-corps were owned by individuals born outside the U.S. and between 75.3 and 77.6 percent were owned by people born in the U.S.. About 3 (3.1 to 3.2) percent of firms were equally owned and the remaining were not assigned a place of birth due to lack of availability of place of birth data or neither place of birth category accounts for fifty percent of ownership shares. It is important to note that there is not a lot of variation in the distribution of firms owned by individuals born outside the U.S. depending on which method is

used, other than a slightly higher percentage in the missing category when we impose the ten percent rule. As with the firm-owner tables, partnerships have a lower percent of firms that owned by individuals born outside the U.S. (12.4 to 12.8 percent) compared to S-corps (24.1 to 24.2 percent). We also find a greater percent of cases with unassigned place of birth among partnerships compared to S-corps.

In Table 41 we show firm-level citizenship distribution. Once again we do not see a great deal of variation depending on method used to assign citizenship. For partnerships, we find that between 5.2 and 5.3 percent of firms are non U.S. citizen-owned, 84.0 to 87.0 percent U.S. citizen-owned, and 2.9 to 3.0 percent equally owned, and for the remainder of partnerships we were not able to assign citizenship status. For S-corps, we find that about 12.9 percent are non U.S. citizen-owned, 82.2 to 82.3 percent were U.S. citizen-owned, and 1.5 to 1.6 percent were equally owned. Once again we find a higher percentage of partnerships with unassigned citizenship status compared to S-corps.

The SBO did not produce estimates of place of birth or citizenship at firm level because these demographic characteristics were not imputed if missing,[80] so no comparison is possible.

E. Age

We also consider the ages of owners of nonemployer firms. Informed by research on age and entrepreneurship (Haltiwanger, et al. 2013; Azulay et al. 2018), we use three age categories: under 30, 30 to 54, and 55 and older. Table 42 reports the age distribution of owner-firm pairs by legal form of organization. Most sole proprietorships are owned by individuals between 30 and 54

[80] In SBO's terminology, they were not "core" demographics.

years of age, with nearly 54 percent falling into this category. Nearly 34 percent are owned by individuals who are at least 55 years old, while about 14 percent are owned by people under 30. Owners of sole proprietorships are 46.6 years old on average.

Owners of partnerships are older than owners of sole proprietorships, averaging 54.6 years of age. This is also reflected in their distribution across age categories, as just over half of partnership owners are 55 or older, while less than five percent are under 30. Owners of S-corps look more similar to partnership owners than sole proprietors, with an average age of 53.2 and just over 46 percent at least 55 years old.

When nonemployer firm owners are counted only once regardless of which type(s) of firm they own, as reported in Table 43, the age distribution again resembles the sole proprietorship distribution, since that is by far the largest group of owners. Their average age is 47.9.

At the firm level, we use the same age categories as at the individual level; we do not use an equally owned category. Table 44 shows the distribution of firm ownership across age categories. Results from the baseline/SBO approach indicate that more than 43 percent of partnerships are majority-owned by people who are 55 years of age or older, and nearly 37 percent are between the ages of 30 and 54. Strikingly, it is not uncommon for firms to fail to reach majority ownership within these age categories; nearly 18 percent of partnerships are not assigned to any of these groups. S-corps are majority owned by people age 30 to 54 nearly 49 percent of the time, and people 55 or older nearly 43 percent of the time. Only about 5 percent of S-corps are not assignable to age categories.

Table 44 also shows firm-level results from the three alternative approaches of assigning firms to demographic categories. Rates of non-assignment are higher across the board, but the

implications of the alternative assignment rules are similar. S-corps are little affected, partnerships see small increases in assignment rates, and the newly assignable firms are allocated mainly to the dominant groups under the baseline approach.

As with place of birth and citizenship, the SBO did not produce estimates of age at firm level, so no comparison is possible.

VII. Challenges and Limitations

In the Methodology section, we described important methodological challenges and how we plan to address them. These included the assignment of demographic characteristics at the firm level, tiered entities, beginning and end-of-year ownership shares, and C-corporations. Here we expand on the discussion of challenges and limitations by providing an overview of issues related to potential non-sampling error in AR and census data sources (e.g., coverage and bias issues), and also of issues regarding data agreements and delivery schedules.

In general, AR data may contain measurement error because of issues such as coverage problems (e.g., the data source may not cover certain populations as well as others), linking or matching issues which may lead to bias problems, conceptual and timing misalignments, reporting errors, missing items or records, etc.. It is, thus, important to have an understanding of how a particular AR or census source may be impacted by these issues.

The primary data sources NES-D uses are well-researched and of high quality. The Census Bureau has done extensive research over the years on decennial, ACS and Census Numident data as well as on linkage issues to identify coverage/underrepresentation and bias. For instance, studies show that very young children, racial and ethnic minorities, low income persons, immigrants not

yet fully integrated in the economy, and mobile persons are missed at higher rates in decennial data. These groups are generally referred to as "hard-to-count" populations. Nevertheless because NES-D's nonemployer universe is largely well integrated into the economy and well represented in tax data, it is unlikely they will be underrepresented to the same extent as the general population in decennial and ACS data. Immigrants entering the U.S. between census years 2000 and 2010 will not be captured in decennial data, and while the ACS will help fill-in some of that gap, race and Hispanic origin data will be missing for most of that immigrant group. In fact, an initial exploration shows that approximately 65 percent of nonemployers (successfully linked to the Numident) with missing race and Hispanic origin information are born outside the U.S.. With that in mind, we are exploring the possibility of using a new Census Bureau algorithm that uses Numident's place of birth information to impute race and ethnicity. Developments on this work will be discussed in a future paper. Our research also showed that older and healthier veterans are underrepresented in VA's AR data.

Meanwhile, by definition, the Census Numident contains only people with SSNs. This seems to impact, but marginally, NES-D nonemployer population since according to our results the vast majority (approximately 99 percent) of nonemployers with PIKs can be linked to the Census Numident. Issues regarding coverage gaps for older and healthier veterans in VA's AR file were fully discussed in the Data section and Appendix.

Other studies have looked at agreement rates in race, ethnicity, sex, age and place of birth values across AR and census records sources (decennial and ACS), and also between all these

sources and reported survey responses.[81] Findings show high agreement rates between Census Numident demographic information and census records, and also between AR and reported responses in survey and decennial data, providing evidence of the suitability of the demographic data sources employed in NES-D to direct replace demographic information in surveys.[82] In addition, AR data on demographic characteristics tend to be less noisy and/or less subject to misreporting errors than magnitude data such as income, or time-varying and/or unobservable information. Still, race and Hispanic origin data in particular are not impervious to misreporting or measurement error. Studies find that agreement rates are lower for small size populations (i.e., AIAN, NHPI, multiracial) relative to other race groups, and for Hispanics relative to non-Hispanics. Research shows that this can be largely attributed to the fact that racial fluidity is more prevalent among these populations.[83]

The Census Bureau has also conducted studies examining biases arising from linking or PIK assignment.[84] This research indicates that certain groups have lower PIK rates than others; namely, younger children, minorities, unemployed and lower-income individuals, and immigrants. Again, these groups tend to be less well represented in tax data since the PIK assignment algorithm relies to a good extent on tax data. Since NES-D's nonemployer universe is well represented in tax data, it will not be as impacted by PIK assignment bias.

[81] See, for instance, Bhaskar (2016), Ennis (2016), Luque (2016), Noon (2016), Rastogi & O'Hara (2012), and Bhaskar et al. (2014)
[82] Also, the 2012 SBO used decennial and ACS data to direct replace and/or impute race, Hispanic origin, sex and veteran status for nonemployer soleproprietors.
[83] Racial fluidity refers to the idea that an individual can be observed as having different races over time or across data sources. See, for instance, Ennis et al. (2015), Liebler et al. (2014) in https://www.census.gov/about/adrm/linkage/projects/socio-econ-demo/race-fluidity.html.
[84] See Bond et al. (2014): https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-08.html .

Non-sampling error due to conceptual misalignments between AR and survey data is well illustrated by the veteran status issue we encountered with VA's AR data. This issue is thoroughly discussed in the Data section and Appendix of this paper.

NES-D's feasibility depends on data agreements of the Census Bureau with the IRS, the Social Security Administration and the Department of Veterans Affairs. Currently, agreements are in place or imminent, but these are subject to change due to unforeseeable circumstances in the future.

Regarding deliveries of tax and Numident data, we do not foresee any unusual issues interfering with the timely delivery of those data. However, timeliness and frequency could be a problem with AR data from VA. As explained earlier, VA's file is a compendium of multiple data sources, and VA's ability to deliver the file on a timely basis depends on the timeliness of other data sources. If annual timely deliveries of VA's data are not possible, NES-D might have to provide estimates on veteran-owned businesses on a less timely and frequent basis (e.g., bi-annually). As discussed in the Data section and Appendix, we are currently looking into Department of Defense data as an alternative source of veteran status information.

Finally, disclosure avoidance rules are evolving and becoming more restrictive. At this point, it is not clear how this will impact official statistics – including NES-D.

VIII. Conclusion and Next Steps

The use of administrative records in surveys and statistics has become more commonplace in the last several years. However, the use of individual-level AR in business surveys has perhaps been under-utilized for varying reasons, including those related to regulatory limitations on how AR

data can be used in surveys. In this context, the creation of the NES-D represents the value of leveraging existing AR and Census Bureau records in business statistics. NES-D will produce quality, more frequent and timely estimates of business demographics with no added respondent burden, and lower imputation rates and costs – issues that have increasingly plagued surveys in the last decade. This set of benefits is particularly important since they address the needs of stakeholders for reliable estimates that are more timely and frequent.

NES-D embodies an innovative approach to producing business statistics, but also and importantly, it is well grounded in a body of proven administrative records research that shows the quality and suitability of the data sources employed in NES-D to direct replace demographic information in household and business surveys. Also thanks to this research we have a good understanding of potential nonsampling errors in our sources of demographic data, such as the aforementioned issues related to coverage, conceptual misalignments, biases in PIK assignment or misreporting.

NES-D is in its nascent stage, and although there are challenges along the way, the initial findings are very promising. As our results show, demographic information can be found in AR and census data for the overwhelming majority of nonemployers. During the next year we plan to address imputation of missing demographics including those of C-corporations, test the longitudinal consistency of our estimates, and develop estimates at detailed industry and geography levels.[85] Estimates will include counts and percentages of nonemployer businesses and their owners, and also receipts estimates. In the third year, this work will transition into the

[85] The work will also involve ensuring compliance with evolving disclosure avoidance rules.

production phase. The goal is to release a beta version of NES-D in 2020 with the 2017 nonemployer vintage.[86]

As discussed in the Challenges section, there are and will be challenges along the way. We will work to meet these challenges, so that NES-D can fulfill the goals of present-day official statistics. Stakeholders in particular, and economic agents in general, are more reliant on data than ever before. To be useful, these data have to be accurate, timely, frequent, consistent, credible and transparent. It is our goal to work to have NES-D fulfill these criteria.

[86] Pending all required reviews.

References

Azulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda. 2018. “Age and High-Growth Entrepreneurship.” Center for Administrative Records Research and Applications Working Paper # 2018-03, U.S. Census Bureau.

Bhaskar, Renuka. 2016. “Evaluation of Nativity Data in the 2012 Survey of Business Owners.” Center for Administrative Records Research and Applications Memorandum #2016-15, U.S. Census Bureau.

Bond, Brittany, J. David Brown, Adela Luque, and Amy O’Hara, 2014. "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey," CARRA Working Papers 2014-08, Center for Economic Studies, U.S. Census Bureau.

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi. 2018. “Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census.” Center for Economic Studies Working Paper 18-38, U.S. Census Bureau.

Ennis, Sharon. 2016 “Evaluation of Race and Ethnicity in the 2012 Survey of Business Owners.” Center for Administrative Records Research and Applications Memorandum #2016-16, U.S. Census Bureau.

Ennis, Sharon R., Sonya R. Porter, James M. Noon, and Ellen Zapata. 2015. "When Race and Hispanic Origin Reporting are Discrepant Across Administrative Records and Third Party Sources: Exploring Methods to Assign Responses," CARRA Working Papers 2015-08, Center for Economic Studies, U.S. Census Bureau.

Goldschlag, Nathan, J. Daniel Kim, and Kristin McCue. 2017. “Just Passing Through: Characterizing U.S. Pass-Through Business Owners.” Center for Economic Studies Working Paper 17-69, U.S. Census Bureau.

Foster, Lucia and Patrice Norman. 2018. “The Annual Survey of Entrepreneurs.” *Journal of Economic and Social Measurement* 42 (3-4), 199-224.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. “Who Creates Jobs? Small Versus Large Versus Young.” *Review of Economics and Statistics* 95(2): 347-361.

Liebler, Carolyn A., Sonya R. Porter, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis. 2017. "America’s Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census," Demography, Springer; Population Association of America (PAA), vol. 54(1), pages 259-284, February.

Luque, Adela. 2016. “Evaluation of PIK Coverage in the 2012 Survey of Business Owners.” Center for Administrative Records Research and Applications Memorandum #2016-14, U.S. Census Bureau.

Luque, Adela, and Renuka Bhaskar. 2014. "2010 American Community Survey Match Study," CARRA Working Papers 2014-03, Center for Economic Studies, U.S. Census Bureau.

Noon, James. 2016. ”Evaluation of Veteran Status Data in the 2012 Survey of Business Owners.” Center for Administrative Records Research and Applications Memorandum #2016-17, U.S. Census Bureau.

Rastogi, Sonya and Amy O’Hara. 2012. “Census Match Study,” 2010 Census Program for Evaluations and Experiments, 2010 Planning Memoranda Series No. 247, U.S. Census Bureau.

Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications (CARRA) Record Linkage Software." Center for Administrative Records Research and Applications Working Paper #2014-01, U.S. Census Bureau.

Tables

Table 1: Sole Proprietorship PIK Availability

	Number	Column Percent
Total	21,020,000	100.00%
Valid PIKs	21,000,000	99.90%
Invalid PIKs	20,000	0.10%

Source: 2015 Nonemployer database.
Note: This table reports the number of valid and invalid PIKs available for owners of sole proprietorships. All values in this and all subsequent tables and figures have been rounded to four significant digits as part of the disclosure avoidance protocol. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000 and above 10,000,000 to the nearest 10,000.

Table 2: K-1 Match by Year of K-1 Data and Legal Form of Organization

	Partnerships		S-Corporations	
	Number	Column Percent	Number	Column Percent
All firms (matched & un-matched)	1,804,000	100%	1,124,000	100%
2014 K-1 data	3,800	0.21%	3,600	0.32%
2015 K-1 data	1,781,000	98.73%	1,102,000	98.04%
2016 K-1 data	6,800	0.38%	7,900	0.70%
Not matched	12,500	0.69%	10,500	0.93%

Source: 2015 Nonemployer database and K-1 data, 2014-2016.
Note: This table reports the number of firms that match to K-1 data, by legal form of organization and the tax year for which K-1 data were matched. Matching was attempted first using 2015 data, then 2016, then 2014. Once a firm matched to a year of K-1 data, it was not included in subsequent merges.

Table 3: Number of Owners by Legal Form of Organization

Number of identified owners (either person or firm owner)	Partnerships owned by either person or firm owner		S-corps owned by either person or firm owner	
	Number	Column Percent	Number	Column Percent
0	0	0.00%	0	0.00%
1	42,500	2.37%	779,000	69.96%
2	1,092,000	61.00%	260,000	23.35%
3	288,000	16.09%	38,000	3.41%
4	153,000	8.55%	18,000	1.62%
5	69,000	3.85%	6,800	0.61%
6	39,000	2.18%	3,800	0.34%
7	22,000	1.23%	2,000	0.18%
8	15,000	0.84%	1,400	0.13%
9	10,500	0.59%	950	0.09%
10	8,300	0.46%	650	0.06%
11	6,500	0.36%	500	0.04%
12	5,300	0.30%	400	0.04%
13	4,000	0.22%	300	0.03%
14	3,500	0.20%	250	0.02%
15+	31,500	1.76%	1,400	0.13%

Note: This table reports the distribution of number of owners for partnerships & S-corps that have been identified in K-1 data. For instance, over 1 million partnerships have 2 owners. Owners include individuals or firms listed in the K-1 data, regardless of ownerships shares. Note that for some firms K-1 data do not identify all owners (i.e., 100 percent firm ownership). See Table 4 below.

Table 4: Owner Coverage of Partnerships and S-corporations, at Firm-Owner Level and Firm Level

	All Firms	Firms Matched to K-1s		Firm-Owner Pairs		Firm-level coverage (firms for which we have identified all owners, narrow definition)		Firm-level coverage (firms for which we have identified all owners, broad definition)	
	Number	Number	Row %	Firm-PIK Pairs	Firm-EIN Pairs	Number	Row Percent (as % of all)	Number	Row Percent (as % of all)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Partnerships	1,804,000	1,791,000	99.28%	5,557,000	1,130,000	1,391,000	77.11%	1,605,000	88.97%
S-Corps	1,124,000	1,113,000	99.02%	1,603,000	29,500	1,076,000	95.73%	1,097,000	97.60%

Source: 2015 Nonemployer database and 2015 K-1 data.

Note: This table presents information on the rates at which partnerships and S-corps can be linked to Form K-1 and the extent to which this linkage identifies all owners of these firms. Firm-owner pairs are reported separately for owners that are people (Firm-PIK pairs) and owners that are other firms (Firm-EIN pairs). In the firm-level coverage columns, the narrow definition of all owners identified includes firms for which the sum of all available ownership shares is either exactly equal to 1 or exactly equal to 100 (some firms report ownership shares as decimals, while others report using whole numbers). The broad definition includes all firms for which the sum of all ownership shares reported is between 0.99 and 1.01 or between 99 and 101 (inclusive in both cases). Ownership shares are obtained from Form K-1. Reported ownership shares from the beginning of the year are used (end of year ownership shares are much less likely to be reported than beginning of the year ownership shares).

Table 5: Nonemployer Tiered Entities

	Firms		Firms Matched to K-1s		Firms Owned Only by People			Firms At Least Partially Owned by Other Firms (Tiered Entities)		
	Number	Column Percent	Number	Row Percent	Number	As percent of Total (row %)	As percent of Matched (row %)	Number	As percent of Total (row %)	As percent of Matched (row %)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total	2,928,000	100.00%	2,905,000	99.21%	2,461,000	84.05%	84.72%	444,000	15.16%	15.28%
Partnerships	1,804,000	61.61%	1,791,000	99.28%	1,367,000	75.78%	76.33%	425,000	23.56%	23.73%
S-corps	1,124,000	38.39%	1,113,000	99.02%	1,094,000	97.33%	98.29%	19,000	1.69%	1.71%

Source: 2015 Nonemployer database and K-1 data, 2014-2016.

Note: This table presents information on the rates at which partnerships and S-corps can be linked to Form K-1 and the prevalence of "tiered entities," firms owned at least partially by other firms.

Table 6: Number of Owners of Nonemployer Firms by Legal Form of Organization Owned

	Total	
	Number	Column Percent
	(1)	(2)
All Groups	24,500,000	100.00%
Sole Proprietorship (SP)	19,220,000	78.45%
Partnership (P)	3,201,000	13.07%
S-corp (S)	1,007,000	4.11%
SP and P	602,000	2.46%
SP and S	265,000	1.08%
P and S	159,000	0.65%
SP, P, and S	41,500	0.17%

Source: 2015 Nonemployer database.

Note: This table reports owner-level information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category

Table 7: Availability of Sex and Age AR, Firm-owner Level, by Legal Form of Organization

	Firm-PIK	Match to Numident		Availability of Sex				Availability of Age	
	Number	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Excluding "Unknown"	As % of firm-PIK pairs	Number	As % of firm-PIK pairs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total	28,160,000	27,880,000	99.01%	27,880,000	99.01%	27,880,000	99.01%	27,880,000	99.01%
Sole Proprietors	21,000,000	20,880,000	99.43%	20,880,000	99.43%	20,870,000	99.38%	20,880,000	99.43%
Partnerships	5,557,000	5,413,000	97.41%	5,413,000	97.41%	5,412,000	97.39%	5,413,000	97.41%
S-Corps	1,603,000	1,594,000	99.44%	1,594,000	99.44%	1,594,000	99.44%	1,594,000	99.44%

Source: 2015 Nonemployer database and Census Numident.

Note: This table presents the availability of sex and age information from the Census Numident by firm type. For partnerships and S-corps, only firm-person owner pairs are considered (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information).

Table 8: Availability of Citizenship & Place of Birth AR, Firm-owner level, by Legal Form of Organization

2015 Nonemployers Database	Firm-PIK	Match to Numident		Availability of Place of Birth		Availability of Citizenship Status			
						Option A Assumes all individuals missing citizenship data are U.S. citizens		Option B Assumes individuals missing citizenship data are U.S. citizens unless place of birth indicates they were born outside the United States	
	Number	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total	28,160,000	27,880,000	99.01	27,880,000	99.01	27,880,000	99.01	27,480,000	97.59
Sole Proprietors	21,000,000	20,880,000	99.43	20,870,000	99.38	20,880,000	99.43	20,660,000	98.38
Partnerships	5,557,000	5,413,000	97.41	5,413,000	97.41	5,413,000	97.41	5,273,000	94.89
S-Corps	1,603,000	1,594,000	99.44	1,594,000	99.44	1,594,000	99.44	1,551,000	96.76

Source: 2015 Nonemployer data and 2015 Numident

Note: This table presents the availability of place of birth and citizenship status information (using the two options for dealing with missing information described above) from the Numident by firm type. For partnerships and S-corps, only firm-person owner pairs are considered (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information).

Table 9: Availability of Previous Census Records (PCR) Race & Hispanic Origin Data, Firm-owner Level, by Legal Form of Organization

	Firm-PIK Pairs	Owner Match to 2015 PCR File		Availability of Race Response not Including Some Other Race		Availability of Some Other Race Response		Availability of Hispanic Origin Response	
	Number	Number	As % of Firm-PIK Pairs	Number	As % of Firm-PIK Pairs	Number	As % of Firm-PIK Pairs	Number	As % of Firm-PIK Pairs
	(1)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Total	28,160,000	26,330,000	93.50	25,320,000	89.91	931,000	3.31	26,330,000	93.50
Sole Proprietorships (tax Form 1040)	21,000,000	19,520,000	92.95	18,580,000	88.48	864,000	4.11	19,520,000	92.95
Partnerships	5,557,000	5,275,000	94.93	5,227,000	94.06	41,000	0.73	5,275,000	94.93
S-corps	1,603,000	1,536,000	95.82	1,507,000	94.01	26,000	1.62	1,536,000	95.82

Source: 2015 Nonemployer data and 2015 Previous Census Records (PCR) File.

Table 10: Number of Demographic Characteristics Missing

Number of Demographic Characteristics	Number Missing	Percent Missing
0	22,430,000	91.55%
1	386,000	1.58%
2	1,428,000	5.83%
3	14,500	0.06%
4	21,500	0.09%
5	100	0.00%
6	223,000	0.91%

Source: 2015 Nonemployer database, Census Numident, decennial census, American Community Survey, and USVets data.
Note: There are 24,500,000 unique owners of sole proprietorships, partnerships, and S-corporations in the data

Table 11: Combination of Demographic Characteristics Missing

Demographic Characteristic	Number Missing	Percent Missing
None	22,430,000	91.55%
Only Numident Variable(s)	327,000	1.33%
Only Race	80,000	0.33%
Only Hispanic Origin	0	0.00%
Only Numident Variable(s) and Race	400	0.00%
Only Numident Variable(s) and Hispanic Origin	0	0.00%
Only Race and Hispanic Origin	1,427,000	5.82%
Numident Variable(s), Race, and Hispanic Origin	238,000	0.97%

Source: 2015 Nonemployer database, Census Numident, decennial census, American Community Survey, and USVets data.
Note: There are 24,500,000 unique owners of sole proprietorships, partnerships, and S-corporations in the data. Race and Hispanic-origin are drawn from the decennial census and American Community Survey. Age, sex, citizenship, and place of birth are drawn from the Numident. To reduce dimensionality, these characteristics are collapsed into a single variable that is equal to one if at least one of them is missing. Veteran status is excluded from this analysis because it is never missing.

Table 12: Owner-Level Demographic Characteristic Missingness, by Characteristic

Demographic Characteristic	Number Missing	Percent Missing
Race	1,745,000	7.12%
Hispanic Origin	1,665,000	6.80%
Any Numident Variable	247,000	1.01%
Age	244,000	1.00%
Sex	247,000	1.01%
Citizenship	557,000	2.27%
Place of birth	251,000	1.02%
Veteran Status	0	0.00%

Source: 2015 Nonemployer database, Census Numident, decennial census, American Community Survey, and USVets data
Note: There are 24,500,000 unique owners of sole proprietorships, partnerships, and S-corporations in the data. Race and Hispanic-origin are drawn from the decennial census and American Community Survey. Age, sex, citizenship, and place of birth are drawn from the Numident. Veteran status is drawn from USVets data. Veteran status is never missing because failure to match to the veteran data indicates that one is not a veteran.

Table 13: Types of Firm-Level Demographic Missingness by Legal Form of Organization

	Sole Props	Partnerships	S-Corps
All Firms	21,020,000	1,804,000	1,124,000
Firms with No Owners Identified			
Number	20,000	12,500	10,500
Percent of All Firms	0.10%	0.69%	0.93%
Firms with Owners Identified			
Number	21,000,000	1,791,500	1,113,500
Percent of All Firms	99.90%	99.31%	99.07%
Some Owners Missing Some Demographics			
Number		211,000	104,000
Percent of All Firms		11.70%	9.25%
Some Owners Missing All Demographics			
Number		21,500	4,000
Percent of All Firms		1.19%	0.36%
All Owners Missing Some Demographics			
Number	1,786,000	44,500	71,000
Percent of All Firms	8.50%	2.47%	6.32%
All Owners Missing All Demographics			
Number	115,000	10,500	3,000
Percent of All Firms	0.55%	0.58%	0.27%

Source: 2015 Nonemployer database, Census Numident, decennial census, American Community Survey, and USVets data.

Note: This table reports the degree to which firms are missing owner demographic information. Categorization is based only on person owners (firm owners do not figure into the categorization).

Table 14: Sex Distribution, Owner-Firm Level, by Legal Form of Organization

	All Firm-Owner Pairs		Sole Proprietorships		Partnerships		S-Corps	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	28,160,000	100%	21,000,000	100%	5,557,000	100%	1,603,000	100%
Male	16,310,000	57.92%	11,520,000	54.86%	3,705,000	66.67%	1,086,000	67.75%
Female	11,570,000	41.09%	9,355,000	44.55%	1,707,000	30.72%	507,000	31.63%
Unknown	3,200	0.01%	2,000	0.01%	1,000	0.02%	250	0.02%
Missing	279,000	0.99%	126,000	0.60%	144,000	2.59%	8,600	0.54%

Source: 2015 Nonemployer database, 2014-2016 K-1 data, and Census Numident.

Note: This table presents firm-owner level sex information by firm type. Only person owners are included in this table (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information). Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories. The "Unknown" category corresponds to a sex code on the Numident. Missing indicates that an individual did not match to the Numident.

Table 15: Sex Distribution, Owner Level, by Legal Form(s) of Organization Owned

	Total		Male		Female		Unknown		Missing	
	Number	Column Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All Groups	24,500,000	100%	13,710,000	55.96%	10,540,000	43.02%	2,600	0.01%	244,000	1.00%
Sole Proprietorship (SP)	19,220,000	78.45%	10,350,000	53.85%	8,745,000	45.50%	1,800	0.01%	123,000	0.64%
Partnership (P)	3,201,000	13.07%	1,948,000	60.86%	1,140,000	35.61%	550	0.02%	113,000	3.53%
S-Corp (S)	1,007,000	4.11%	646,000	64.15%	353,000	35.05%	150	0.01%	7,800	0.77%
SP and P	602,000	2.46%	427,000	70.93%	174,000	28.90%	70	0.01%	500	0.08%
SP and S	265,000	1.08%	187,000	70.57%	77,000	29.06%	30	0.01%	450	0.17%
P and S	159,000	0.65%	120,000	75.47%	39,500	24.84%	20	0.01%	100	0.06%
SP, P, and S	41,500	0.17%	34,000	81.93%	7,500	18.07%	<15	(D)	<15	(D)

Source: 2015 Nonemployer database, 2014-2016 K-1 data, and Census Numident.

Note: This table reports owner-level sex information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. The "Unknown" category corresponds to a sex code listed on the Numident, while missing indicates that an individual was not matched to the Numident.

N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table 16: Distribution of Number of Firms Owned

	Nonemployer Firms Owned	
Owners	Number	Column Percent
1	22,060,000	90.04%
2	1,892,000	7.72%
3	328,000	1.34%
4	98,000	0.40%
5	45,000	0.18%
6	24,500	0.10%
7	14,500	0.06%
8	9,500	0.04%
9	6,400	0.03%
10	4,600	0.02%
11	3,400	0.01%
12	2,600	0.01%
13	2,100	0.01%
14	1,500	0.01%
15+	7,600	0.03%

Note: This table reports the distribution of the number of firms owned by individual owners. The Nonemployer Firms Owned column include owners of any type or types of nonemployer firms (i.e., soleproprietorships, partnerships & S-corps). For instance, over 22 million nonemployer firms have one owner, and close to 2 million firms have two owners.

Table 17: Firm Ownership Sex Distribution, by Legal Form of Organization and Method of Assignment

	Total	Male		Female		Equal		Unknown		Not Assigned	
	Number	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
All											
Four Owners, 10% Rule (SBO)	2,787,000	1,576,000	56.55%	515,000	18.48%	573,000	20.56%	200	0.01%	123,000	4.41%
Four Owners, No 10% Rule	2,787,000	1,623,000	58.23%	527,000	18.91%	583,000	20.92%	200	0.01%	54,000	1.94%
All Owners, 10% Rule	2,787,000	1,593,000	57.16%	521,000	18.69%	569,000	20.42%	200	0.01%	103,000	3.70%
All Owners, No 10% Rule	2,787,000	1,630,000	58.49%	530,000	19.02%	573,000	20.56%	200	0.01%	54,000	1.94%
Partnerships											
Four Owners, 10% Rule (SBO)	1,681,000	873,000	51.93%	258,000	15.35%	438,000	26.06%	100	0.01%	112,000	6.66%
Four Owners, No 10% Rule	1,681,000	919,000	54.67%	270,000	16.06%	448,000	26.65%	100	0.01%	44,000	2.62%
All Owners, 10% Rule	1,681,000	890,000	52.94%	264,000	15.70%	435,000	25.88%	100	0.01%	91,510	5.44%
All Owners, No 10% Rule	1,681,000	925,000	55.03%	272,000	16.18%	439,000	26.12%	100	0.01%	43,500	2.59%
S-Corps											
Four Owners, 10% Rule (SBO)	1,106,000	703,000	63.56%	257,000	23.24%	135,000	12.21%	100	0.01%	11,500	1.04%
Four Owners, No 10% Rule	1,106,000	704,000	63.65%	257,000	23.24%	135,000	12.21%	100	0.01%	10,000	0.90%
All Owners, 10% Rule	1,106,000	703,000	63.56%	257,000	23.24%	134,000	12.12%	100	0.01%	11,500	1.04%
All Owners, No 10% Rule	1,106,000	704,000	63.65%	258,000	23.33%	134,000	12.12%	100	0.01%	10,000	0.90%

Source: 2015 Nonemployer database, 2014-2016 K-1 data, and 2015 Census Numident.

Note: This table reports firm level ownerships by sex for partnerships and s-corps. Owners are identified from Schedule K-1. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Firms are assigned to groups based on sex using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. A firm is assigned to a sex group if members of that group account for more than 50 percent of ownership reported by the indicated number of owners. Firms in which ownership is split 50-50 between men and women are assigned to the "Equal" group. Firms in which no sex group accounts for more than 50 percent of ownership are not assigned to a given sex group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned to a sex category; where "No 10% Rule" is indicated, those firms are assigned to a sex category if possible. The "Unknown" category corresponds to a sex code obtained from the Numident. Ownership shares belonging to other firms do not figure into this process.

Table 18: Comparison to SBO, Sex, Firm Level

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO*		Total - 2015 Adrec, 10% Rule		Total - 2015 Adrec, No 10% Rule	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total	17,348,744	100%	21,357,346	100%	22,201,901	100%	23,790,000	100%	23,790,000	100%
Male	9,659,064	55.68%	10,670,479	49.96%	11,509,025	51.84%	13,100,000	55.07%	13,140,000	55.23%
Female	5,572,602	32.12%	6,882,453	32.23%	8,842,742	39.83%	9,870,000	41.49%	9,883,000	41.54%
Equal	1,975,399	11.39%	3,551,960	16.63%	1,691,409	7.62%	573,000	2.41%	583,000	2.45%
Not Assigned	141,679	0.82%	252,454	1.18%	158,725	0.71%	252,000	1.06%	183,000	0.77%

Source: Survey of Business Owners, 2002, 2007, and 2012; 2015 Nonemployer database, 2014-2016 K-1 data, and 2015 Census Numident.

Note: This table reports firm-level ownership by sex based on data obtained from SBO published tables (first six columns) and 2015 administrative records (last four columns). The SBO columns include owners of all firms without paid employees. The administrative records columns include owners of nonemployer sole proprietorships, partnerships, and S-corps. In the administrative records columns, firm ownership is assigned to partnerships and S-corps using the four person owners with the largest ownership shares. In columns seven and eight, sex is not assigned to firms in which the person who owns the largest share owns less than ten percent of the firm; in columns nine and ten, sex is assigned to those firms if otherwise possible. These columns report firms assigned to "unknown" sex in the same category as firms that could not be assigned a sex.

*See https://www.census.gov/data/tables/2012/econ/sbo/2012-sbo-company-summary.html for SBO tables showing relative standard errors.

Table 19: Race Distribution, Firm-owner level, by Legal Form of Organization

	2015 Previous Census Records Data	
	Number	Column %
Total	28,160,000	100.00
White	20,890,000	74.18
Black	2,363,000	8.39
American Indian Alaska Native	50,000	0.18
Asian	1,531,000	5.44
Native Hawaiian or Pacific Islander	24,500	0.09
Multiple Races	456,000	1.62
Missing	2,847,000	10.11
Sole Proprietorships (tax Form 1040)		
Total	21,000,000	100.00
White	14,840,000	70.66
Black	2,221,000	10.58
American Indian Alaska Native	44,500	0.21
Asian	1,073,000	5.11
Native Hawaiian or Pacific Islander	21,000	0.10
Multiple Races	381,000	1.81
Missing	2,421,000	11.53
Partnerships & S-corps (tax form K-1)		
Total	7,160,000	100.00
White	6,050,000	84.50
Black	142,000	1.98
American Indian Alaska Native	5,400	0.07
Asian	458,000	6.40
Native Hawaiian or Pacific Islander	3,200	0.05
Multiple Races	75,500	1.06
Missing	426,000	5.95
Partnerships Total	5,557,000	100.00
White	4,738,000	85.26
Black	84,500	1.52
American Indian Alaska Native	3,700	0.07
Asian	345,000	6.20
Native Hawaiian or Pacific Islander	2,300	0.04
Multiple Races	53,500	0.96
Missing	330,000	5.95
S-corps Total	1,603,000	100.00
White	1,312,000	81.84
Black	57,500	3.58
American Indian Alaska Native	1,600	0.10
Asian	114,000	7.10
Native Hawaiian or Pacific Islander	900	0.06
Multiple Races	22,000	1.38
Missing	95,500	5.95

Source: 2015 Nonemployer data, 2015 Previous Census Records file.

Note: Only person owners are included in this table. Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories.

* Some Other Race is not included in the distribution

Table 20: Race Distribution (excluding missing values), Firm-owner level, by LFO

	2015 Previous Census Records Data	
	Number	Column %
Total	25,320,000	100.00
White	20,890,000	82.50
Black	2,363,000	9.33
American Indian Alaska Native	50,000	0.20
Asian	1,531,000	6.05
Native Hawaiian or Pacific Islander	24,500	0.10
Multiple Races	456,000	1.80
Sole Proprietorships		
Total	18,580,000	100.00
White	14,840,000	79.87
Black	2,221,000	11.95
American Indian Alaska Native	44,500	0.24
Asian	1,073,000	5.78
Native Hawaiian or Pacific Islander	21,000	0.11
Multiple Races	381,000	2.05
Partnerships & S-corps		
Total	6,734,000	100.00
White	6,050,000	89.84
Black	142,000	2.11
American Indian Alaska Native	5,400	0.08
Asian	458,000	6.80
Native Hawaiian or Pacific Islander	3,200	0.05
Multiple Races	75,500	1.12
Partnerships		
Total	5,227,000	100.00
White	4,738,000	90.64
Black	84,500	1.62
American Indian Alaska Native	3,700	0.07
Asian	345,000	6.60
Native Hawaiian or Pacific Islander	2,300	0.04
Multiple Races	53,500	1.02
S-corps		
Total	1,507,000	100.00
White	1,312,000	87.06
Black	57,500	3.82
American Indian Alaska Native	1,600	0.11
Asian	114,000	7.56
Native Hawaiian or Pacific Islander	900	0.06
Multiple Races	22,000	1.46

Source: 2015 Nonemployer data, 2015 Previous Census Records file.

Table 21: Hispanic Origin Distribution, Firm-owner level, by Legal Form of Organization

	2015 Previous Census Records Data	
	Number	Column %
Hispanic Origin		
Total	28,160,000	100.00
Hispanic	2,920,000	10.37
Non-Hispanic	23,410,000	83.13
Missing	1,829,000	6.50
Sole Proprietorships		
Total	21,000,000	100.00
Hispanic	2,595,000	12.36
Non-Hispanic	16,930,000	80.62
Missing	1,481,000	7.05
Partnerships & S-corps		
Total	7,160,000	100.00
Hispanic	325,000	4.54
Non-Hispanic	6,487,000	90.60
Missing	349,000	4.87
Partnerships		
Total	5,557,000	100.00
Hispanic	193,000	3.47
Non-Hispanic	5,081,000	91.43
Missing	282,000	5.07
S-corps		
Total	1,603,000	100.00
Hispanic	131,000	8.17
Non-Hispanic	1,405,000	87.65
Missing	66,000	4.12

Source: 2015 Nonemployer data, 2015 Previous Census Records file.

Table 22: Hispanic Origin Distribution (excluding missing values), Firm-owner level, by Legal Form of Organization

	2015 Previous Census Records Data	
	Number	Column %
Hispanic Origin		
Total	26,330,000	100.00
Hispanic	2,920,000	11.09
Non-Hispanic	23,410,000	88.91
Sole Proprietorships		
Total	19,520,000	100.00
Hispanic	2,595,000	13.29
Non-Hispanic	16,930,000	89.73
Partnerships & S-corps		
Total	6,811,000	100.00
Hispanic	325,000	4.77
Non-Hispanic	6,487,000	95.24
Partnerships		
Total	5,275,000	100.00
Hispanic	193,000	3.66
Non-Hispanic	5,081,000	96.32
S-corps		
Total	1,536,000	100.00
Hispanic	131,000	8.53
Non-Hispanic	1,405,000	91.47

Source: 2015 Nonemployer data, 2015 Previous Census Records file, and Race Administrative Records Composite file.

Table 23: Availability of Previous Census Records (PCR) Race & Hispanic Origin Data, Owner level, by Legal Form of Organization

2015 Nonemployers Database	Owners	Owner Match to 2015 PCR File		Availability of Race Response not Including Some Other Race		Availability of Some Other Race Response		Availability of Hispanic Origin Response	
	Number	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs
Total	24,500,000	22,830,000	93.57	21,870,000	89.63	882,000	3.61	22,830,000	93.57
Partnership	3,201,000	3,005,000	93.88	2,971,000	92.81	28,500	0.89	3,005,000	93.88
Partnership and S-corp	159,000	156,000	98.11	155,000	97.48	950	0.60	156,000	98.11
Sole proprietorship	19,220,000	17,830,000	92.77	16,940,000	88.14	823,000	4.28	17,830,000	92.77
Sole proprietorship and Partnership	602,000	587,000	97.51	580,000	96.35	5,700	0.95	587,000	97.51
Sole proprietorship, Partnership, and S-corp	41,500	40,500	97.59	40,000	96.39	250	0.60	40,500	97.59
Sole proprietorship and S-corp	265,000	248,000	93.58	241,000	90.94	6,300	2.38	248,000	93.58
S-corp	1,007,000	967,000	96.03	947,000	94.04	17,500	1.74	967,000	96.03

Source: 2015 Nonemployer data, and 2015 Previous Census Records file.
Note: This table reports owner-level race information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category.

Table 24: Race Distribution, Owner level, by Legal Form of Organization

2015 Nonemployers Database	Total	White		Black		American Indian or Alaska Native		Asian		Native Hawaiian or Pacific Islander		Multiple Races		Missing	
	Number	Number	Row %	Number	Row %	Number	Row %	Number	Row %	Number	Row %	Number	Row %	Number	Row %
Total	24,500,000	17,837,500	72.81	2,238,850	9.14	46,570	0.19	1,309,500	5.34	22,780	0.09	412,550	1.68	2,627,550	10.72
Partnership	3,201,000	2,670,000	83.41	58000	1.81	2600	0.08	207,000	6.47	1600	0.05	32500	1.02	229,000	7.15
Partnership and S-Corp	159,000	140,000	88.05	1,800	1.13	90	0.06	11,000	6.92	60	0.04	1,600	1.01	4,500	2.83
Sole proprietorship	19,220,000	13,440,000	69.93	2,113,000	10.99	42,000	0.22	966,000	5.03	20,000	0.10	353,000	1.84	2,286,000	11.89
Sole proprietorship and Partnership	602,000	525,000	87.21	13,000	2.16	450	0.07	34,000	5.65	300	0.05	6,700	1.11	22,550	3.75
Sole proprietorship, Partnership, and S-Corp	41,500	36,500	87.95	550	1.33	30	0.07	2,500	6.02	20	0.05	450	1.08	1,500	3.61
Sole proprietorship and S-Corp	265,000	206,000	77.74	12,000	4.53	300	0.11	18,000	6.79	200	0.08	4,300	1.62	24,000	9.06
S-Corp	1,007,000	820,000	81.43	40500	4.02	1100	0.11	71,000	7.05	600	0.06	14,000	1.39	60,000	5.96

Source: 2015 Nonemployer data, 2015 Previous Census Records file.

Note: This table reports owner-level race information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category.

Table 25: Hispanic Origin Distribution, Owner level, by Legal Form of Organization

2015 Nonemployers Database	Total	Hispanic		Non-Hispanic		Missing	
	Number	Number	Percent	Number	Percent	Number	Percent
Total	24,500,000	2,713,100	11.08	20,119,500	82.12	1,667,400	6.80
Partnership	3,201,000	123,000	3.86	2,881,000	90.01	197,000	6.14
Partnership and S-corp	159,000	6,000	3.77	150,000	94.11	3,000	2.12
Sole proprietorship	19,220,000	2,441,000	12.70	15,390,000	80.07	1,389,000	7.22
Sole proprietorship and Partnership	602,000	26,500	4.37	560,000	93.06	15,500	2.56
Sole proprietorship, Partnership, and S-corp	41,500	1,600	3.95	38,500	93.21	1,400	2.84
Sole proprietorship and S-corp	265,000	32,500	12.23	216,000	81.42	16,500	6.36
S-corp	1,007,000	82,500	8.21	884,000	87.79	40,500	4.00

Source: 2015 Nonemployer data, 2015 Previous Census Records file.
Note: This table reports owner-level Hispanic origin information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category.

Table 26: Firm Ownership Race Distribution, by Legal Form of Organization & Method of Assignment

	Total	White		Black		AIAN		Asian		NHPI		Multiple Races		Not Assigned		Unlinkable	
	N	N	Row %	N	Row %	N	Row %	N	Row %	NHPI	Row %	N	Row %	N	Row %	N	Row %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
All																	
Four Owners, 10% Rule	23,790,000	17,090,000	71.84%	2,290,000	9.63%	46,000	0.19%	1,234,000	5.19%	22,000	0.09%	403,000	1.69%	1,152,000	4.84%	1,555,000	6.54%
Four Owners, No 10% Rule	23,790,000	17,130,000	72.01%	2,291,000	9.63%	46,000	0.19%	1,236,000	5.20%	22,000	0.09%	403,000	1.69%	1,102,000	4.63%	1,555,000	6.54%
All Owners, 10% Rule	23,790,000	17,090,000	71.84%	2,290,000	9.63%	46,000	0.19%	1,234,000	5.19%	22,000	0.09%	403,000	1.69%	1,148,000	4.83%	1,555,000	6.54%
All Owners, No 10% Rule	23,790,000	17,140,000	72.05%	2,290,000	9.63%	46,000	0.19%	1,236,000	5.20%	22,000	0.09%	403,000	1.69%	1,101,000	4.63%	1,555,000	6.54%
Sole Proprietorships	21,000,000	14,840,000	70.67%	2,221,000	10.58%	44,500	0.21%	1,073,000	5.11%	21,000	0.10%	381,000	1.81%	940,000	4.48%	1,481,000	7.05%
Partnerships																	
Four Owners, 10% Rule	1,681,000	1,364,000	81.14%	25,500	1.52%	600	0.04%	86,500	5.15%	400	0.02%	8,500	0.51%	166,000	9.88%	29,500	1.75%
Four Owners, No 10% Rule	1,681,000	1,411,000	83.94%	25,500	1.52%	600	0.04%	88,500	5.26%	400	0.02%	8,700	0.52%	116,000	6.90%	29,500	1.75%
All Owners, 10% Rule	1,681,000	1,368,000	81.38%	25,500	1.52%	600	0.04%	86,500	5.15%	400	0.02%	8,500	0.51%	162,000	9.64%	29,500	1.75%
All Owners, No 10% Rule	1,681,000	1,412,000	84.00%	25,500	1.52%	600	0.04%	88,000	5.23%	400	0.02%	8,600	0.51%	115,000	6.84%	29,500	1.75%
S-Corps																	
Four Owners, 10% Rule	1,106,000	881,000	79.66%	43,500	3.93%	1,000	0.09%	75,000	6.78%	500	0.05%	14,000	1.27%	46,000	4.16%	44,500	4.02%
Four Owners, No 10% Rule	1,106,000	882,000	79.75%	43,500	3.93%	1,000	0.09%	75,000	6.78%	500	0.05%	14,000	1.27%	45,000	4.07%	44,500	4.02%
All Owners, 10% Rule	1,106,000	881,000	79.66%	43,500	3.93%	1,000	0.09%	75,000	6.78%	500	0.05%	14,000	1.27%	46,000	4.16%	44,500	4.02%
All Owners, No 10% Rule	1,106,000	883,000	79.84%	43,500	3.93%	1,000	0.09%	75,000	6.78%	500	0.05%	14,000	1.27%	45,000	4.07%	44,500	4.02%

Source: 2015 Nonemployer data, 2015 Previous Census Records file.

Note: This table reports firm level ownerships by race by LFO. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Partnerships and S-corps are assigned to race groups using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. A firm is assigned to a race group if members of that group account for more than 50 percent of ownership reported by the indicated number of owners. Firms in which no race group accounts for more than 50 percent of ownership are not assigned to a given race group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned to a race category; where "No 10% Rule" is indicated, those firms are assigned to a race category if possible.

Table 27: Firm Ownership Hispanic Origin Distribution, by Legal Form of Organization & Method of Assignment

	Total	Hispanic		Non-Hispanic		Equal		Not Assigned		Unlinkable	
	Number	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
All											
Four Owners, 10% Rule (SBO)	23,790,000	2,743,000	11.53%	19,330,000	81.25%	34,000	0.14%	124,000	0.52%	1,555,000	6.54%
Four Owners, No 10% Rule	23,790,000	2,744,000	11.53%	19,380,000	81.46%	34,500	0.15%	73,500	0.31%	1,555,000	6.54%
All Owners, 10% Rule	23,790,000	2,743,000	11.53%	19,340,000	81.29%	34,000	0.14%	121,000	0.51%	1,555,000	6.54%
All Owners, No 10% Rule	23,790,000	2,744,000	11.53%	19,380,000	81.46%	34,000	0.14%	73,000	0.31%	1,555,000	6.54%
Sole Proprietorships	21,000,000	2,595,000	12.36%	16,930,000	80.62%	0	0.00%	0	0.00%	1,481,000	7.05%
Partnerships											
Four Owners, 10% Rule (SBO)	1,681,000	52,500	3.12%	1,462,000	86.97%	25,500	1.52%	111,000	6.60%	29,500	1.75%
Four Owners, No 10% Rule	1,681,000	53,500	3.18%	1,511,000	89.89%	26,000	1.55%	61,000	3.63%	29,500	1.75%
All Owners, 10% Rule	1,681,000	52,500	3.12%	1,466,000	87.21%	25,500	1.52%	107,000	6.37%	29,500	1.75%
All Owners, No 10% Rule	1,681,000	53,500	3.18%	1,512,000	89.95%	25,500	1.52%	60,500	3.60%	29,500	1.75%
S-corps											
Four Owners, 10% Rule (SBO)	1,106,000	95,500	8.63%	943,000	85.26%	8,600	0.78%	14,000	1.27%	44,500	4.02%
Four Owners, No 10% Rule	1,106,000	96,000	8.68%	945,000	85.44%	8,600	0.78%	12,500	1.13%	44,500	4.02%
All Owners, 10% Rule	1,106,000	95,000	8.59%	944,000	85.35%	8,500	0.77%	13,500	1.22%	44,500	4.02%
All Owners, No 10% Rule	1,106,000	96,000	8.68%	945,000	85.44%	8,500	0.77%	12,500	1.13%	44,500	4.02%

Source: 2015 Nonemployer data, 2015 Previous Census Records data.

Note: This table reports firm level ownerships by Hispanic origin by LFO. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Partnerships and S-corps are assigned to Hispanic origin groups using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. A firm is assigned to a Hispanic origin group if members of that group account for more than 50 percent of ownership reported by the indicated number of owners. Firms in which ownership is split 50-50 between Hispanic and non-Hispanic are assigned to the "Equal" group. Firms in which no Hispanic origin group accounts for more than 50 percent of ownership are not assigned to a given race group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned to a Hispanic origin category; where "No 10% Rule" is indicated, those firms are assigned to a Hispanic origin category if possible.

Table 28: Firm-level Ownership by Race in the SBO & Previous Census Records (PCR) Data

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO*		Total - 2015 PCR, 10% Rule		Total - 2015 PCR, No 10% Rule	
	Number	Column %	Number	Column %	Number	Column %	Number	Column %	Number	Column %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total	17,348,744	100.0	21,357,346	100.0	22,201,901	100.0	23,790,000	100.0	23,790,000	100.0
White	15,187,720	87.54	17,955,403	84.07	17,101,796	77.03	17,090,000	71.84	17,130,000	72.01
Black	1,103,049	6.36	1,815,298	8.50	2,475,266	11.15	2,290,000	9.63	2,291,000	9.63
AIAN	176,889	1.02	213,029	1.00	246,740	1.11	46,000	0.19	46,000	0.19
Asian	784,118	4.52	1,152,134	5.39	1,436,876	6.47	1,234,000	5.19	1,236,000	5.20
NHPI	25,255	0.15	33,536	0.16	50,043	0.23	22,000	0.09	22,000	0.09
Some Other Race	N/A	N/A	67,366	0.32	1,094,974	4.93	N/A	N/A	N/A	N/A
Multiple Races	N/A	N/A	N/A	N/A	N/A	N/A	403,000	1.69	403,000	1.69
Not Assigned	141,679	0.82	252,454	1.18	158,725	0.71	2,707,000	11.38	2,657,000	11.17

Note: SBO percentages may sum greater than 100% because SBO firms can be assigned more than one race. 2015 figures use the four owners rules.

*See https://www.census.gov/data/tables/2012/econ/sbo/2012-sbo-company-summary.html for SBO tables showing relative standard errors.

N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table 29: Firm-level Ownership by Hispanic Origin in the SBO & Previous Census Records (PCR) Data (includes Not-Assigned firms)

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO*		Total - 2015 PCR, 10% Rule		Total - 2015 PCR, No 10% Rule	
	Number	Column %	Number	Column %	Number	Column %	Number	Column %	Number	Column %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total	17,348,744	100.0	21,357,346	100.0	22,201,901	100.0	23,790,000	100.0	23,790,000	100.0
Hispanic	1,373,922	7.92	2,011,417	9.42	3,018,371	13.60	2,743,000	11.53	2,744,000	11.53
Non-Hispanic	15,833,146	91.26	18,897,297	88.48	18,933,591	85.28	19,330,000	81.25	19,380,000	81.46
Equal	0	0.00	196,178	0.92	91,214	0.41	34,000	0.14	34,500	0.15
Not Assigned	141,679	0.82	252,454	1.18	158,725	0.71	1,679,000	7.06	1,628,000	6.84

Note: SBO percentages may sum greater than 100% because SBO firms can be assigned more than one race. 2015 figures use the four owners rules.

*See https://www.census.gov/data/tables/2012/econ/sbo/2012-sbo-company-summary.html for SBO tables showing relative standard errors.

Table 30: Firm-level Ownership by Race in the SBO & Previous Census Records (PCR) Data (excludes missing values)

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO		Total - 2015 PCR, 10% Rule		Total - 2015 PCR, No 10% Rule	
	Number	Column %	Number	Column %	Number	Column %	Number	Column %	Number	Column %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total	17,277,031	100.0	21,236,766	100.0	22,405,695	100.0	21,080,000	100.0	21,128,000	100.0
White	15,187,720	87.91	17,955,403	84.55	17,101,796	76.33	17,090,000	81.07	17,130,000	81.08
Black	1,103,049	6.38	1,815,298	8.55	2,475,266	11.05	2,290,000	10.86	2,291,000	10.84
AIAN	176,889	1.02	213,029	1.00	246,740	1.10	46,000	0.22	46,000	0.22
Asian	784,118	4.54	1,152,134	5.43	1,436,876	6.41	1,234,000	5.85	1,236,000	5.85
NHPI	25,255	0.15	33,536	0.16	50,043	0.22	22,000	0.10	22,000	0.10
Some Other Race	N/A	N/A	67,366	0.32	1,094,974	4.89	N/A	N/A	N/A	N/A
Multiple Races	N/A	N/A	N/A	N/A	N/A	N/A	403,000	1.91	403,000	1.91

Notes: SBO percentages may sum greater than 100% because SBO firms can be assigned more than one race. 2015 figures use the four owners and 10 percent rules.

N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table 31: Firm-level Ownership by Hispanic Origin in the SBO & Previous Census Records (PCR) Data (excludes missing values)

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO		Total - 2015 PCR, 10% Rule		Total - 2015 PCR No 10% Rule	
	Number	Column %	Number	Column %	Number	Column %	Number	Column %	Number	Column %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total	17,207,068	100.0	20,908,714	100.0	21,951,962	100.0	22,080,000	100.0	22,130,000	100.0
Hispanic	1,373,922	7.98	2,011,417	9.62	3,018,371	13.75	2,743,000	12.42	2,744,000	12.40
Non-Hispanic	15,833,146	92.02	18,897,297	90.38	18,933,591	86.25	19,330,000	87.55	19,380,000	87.57
Equal	0	0.00	196,178	0.94	91,214	0.42	34,000	0.15	34,500	0.16

Note: SBO percentages may sum greater than 100% because SBO firms can be assigned more than one race. 2015 figures use the four owners rules.

Table 32: Veteran Status from 2015 VA USVETS Data, Firm-owner level, by Legal Form of Organization

Veteran Status	Firm-PIK pairs		Sole Proprietors		Partnerships		S-Corps	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	28,160,000	100.00%	21,000,000	100.00%	5,557,000	100.00%	1,603,000	100.00%
Veteran	1,766,000	6.27%	1,301,000	6.20%	354,000	6.37%	111,000	6.92%
Non-veteran	26,400,000	93.75%	19,700,000	93.81%	5,203,000	93.63%	1,492,000	93.08%

Source: 2015 Nonemployer data and 2015 VA USVETS.

Note: This table presents firm-owner level veteran status information by firm LFO. Owners of sole proprietorships are identified from Form 1040 (via the Business Register nonemployer extract). Owners of partnerships and S-corps are identified from Schedule K-1. Veteran status information is obtained from the 2015 VA USVETS data. Individuals matched to the VA USVETS data are considered veterans, while those who did not match the VA USVETS data are considered non-veterans. Only person owners are included in this table (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information). Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories.

Table 33: Veteran Status from 2015 VA USVETS Data, Owner Level, by Legal form of Organization

Owner type	Total Number	Veteran		Non-veteran	
		Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)
Total	24,500,000	1,552,000	6.3%	22,950,000	93.7%
Sole Proprietorship	19,220,000	1,193,000	6.2%	18,030,000	93.8%
Partnership	3,201,000	217,000	6.8%	2,984,000	93.2%
S-Corp	1,007,000	76,500	7.6%	931,000	92.5%
Sole Proprietorship and Partnership	602,000	38,000	6.3%	564,000	93.7%
Sole Proprietorship and S-Corp	265,000	16,000	6.0%	249,000	94.0%
Partnership and S-Corp	159,000	9,300	5.9%	150,000	94.3%
Sole Proprietorship, Partnership, and S-Corp	41,500	2,500	6.0%	39,000	94.0%

Source: 2015 Nonemployer data and 2015 VA USVETS.

Note: This table reports owner-level veteran information by LFO of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. Owners of sole proprietorships are identified from Form 1040. Owners of partnerships and S-corps are identified from Schedule K-1. Veteran status information is obtained from the 2015 VA USVETS data. Individuals matched to the VA USVETS data are considered veterans, while those who did not match the VA USVETS data are considered non-veterans.

Table 34: Veteran Status from 2015 VA USVETS Data, Firm Level, by Firm LFO & Method of Assignment

Firm Level Ownership	Total Number	Veteran		Equal		Non-veteran		Missing	
		Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Partnerships and S-Corps									
Four Owners, 10% Rule (SBO)	2,787,000	109,000	3.9%	104,000	3.7%	2,496,000	89.6%	78,000	2.8%
Four Owners, No 10% Rule	2,787,000	111,000	4.0%	105,000	3.8%	2,571,000	92.3%	0	0.0%
All Owners, 10% Rule	2,787,000	108,000	3.9%	103,000	3.7%	2,501,000	89.7%	75,000	2.7%
All Owners, No 10% Rule	2,787,000	109,000	3.9%	104,000	3.7%	2,574,000	92.4%	0	0.0%
Partnerships									
Four Owners, 10% Rule (SBO)	1,681,000	45,000	2.7%	81,500	4.9%	1,480,000	88.0%	74,000	4.4%
Four Owners, No 10% Rule	1,681,000	46,500	2.8%	82,000	4.9%	1,552,000	92.3%	0	0.0%
All Owners, 10% Rule	1,681,000	44,000	2.6%	80,500	4.8%	1,485,000	88.3%	71,000	4.2%
All Owners, No 10% Rule	1,681,000	45,000	2.7%	81,000	4.8%	1,554,000	92.4%	0	0.0%
S-Corps									
Four Owners, 10% Rule (SBO)	1,106,000	64,000	5.8%	23,000	2.1%	1,015,000	91.8%	4,000	0.4%
Four Owners, No 10% Rule	1,106,000	64,000	5.8%	23,000	2.1%	1,019,000	92.1%	0	0.0%
All Owners, 10% Rule	1,106,000	64,000	5.8%	23,000	2.1%	1,016,000	91.9%	4,000	0.4%
All Owners, No 10% Rule	1,106,000	64,000	5.8%	23,000	2.1%	1,019,000	92.1%	0	0.0%
Sole Proprietors									
firm-PIK = firm level	21,000,000	1,301,000	6.2%	0	0.0%	19,700,000	93.8%	0	0.0%

Source: 2015 Nonemployer data and 2015 VA USVETS.

Note: This table reports firm level ownerships by veteran status for partnerships, S-corps, and sole proprietors. Owners are identified from Schedule K-1. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Firms are assigned to veteran status using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. Depending on the method of assignment, a firm is assigned to a veteran status group if members of that group account for more than 50 percent of ownership reported by either the four largest owners or all owners. Firms in which ownership is split 50-50 between veterans and non-veterans are assigned to the "Equal" group. Firms in which veterans do not account for more than 50 percent of ownership are assigned to non-veteran group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned a veteran status. Ownership shares belonging to other firms do not figure into this process.

Table 35: Firm-level Veteran Status in the SBO & VA USVETS Data

Veteran Status	Total - 2007 SBO*		Total - 2012 SBO**		Total - 2015 VA USVETS		Total - 2015 VA USVETS		Total - 2015 VA USVETS		Total - 2015 VA USVETS	
					4 owners w/ 10%		4 owners w/out 10%		all owners w/ 10%		all owners w/out 10%	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total	21,357,346	100.0%	22,201,902	100.0%	23,790,000	100.0%	23,790,000	100.0%	23,790,000	100.0%	23,790,000	100.0%
Veteran	1,956,259	9.2%	2,079,197	9.4%	1,410,000	5.9%	1,412,000	5.9%	1,409,000	5.9%	1,410,000	5.9%
Equal	948,820	4.4%	412,693	1.9%	104,000	0.4%	105,000	0.4%	103,000	0.4%	104,000	0.4%
Non-veteran	18,199,813	85.2%	19,551,287	88.1%	22,200,000	93.3%	22,270,000	93.6%	22,200,000	93.3%	22,280,000	93.7%
Missing	252,454	1.2%	158,725	0.7%	78,000	0.3%	0	0.0%	75,000	0.3%	0	0.0%

Source: 2015 Nonemployer data and 2015 VA USVETS

Note: This table reports firm-level ownership by veteran status based on data obtained from 2007 and 2012 SBO published tables (left columns) and 2015 VA USVETS data (right columns). The 2007 and 2012 SBO columns include owners of all firms without paid employees. The VA USVETS data columns include owners of nonemployer sole proprietorships, partnerships, and S-corps (excluding C-corps). In the VA USVETS records columns, firm ownership is assigned to partnerships and S-corps using the stated measure.

*2002 SBO does not report veteran status at the firm level. **See https://www.census.gov/data/tables/2012/econ/sbo/2012-sbo-company-summary.html for SBO tables showing relative standard errors.

Table 36: Place of Birth Distribution, Firm-owner Level, by Legal Form of Organization

	All Firm-Owner Pairs		Sole Proprietorships		Partnerships		S-Corps	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	28,160,000	100.00%	21,000,000	100.00%	5,557,000	100.00%	1,603,000	100.00%
Born in the United States*	22,200,000	78.84%	16,340,000	77.81%	4,641,000	83.52%	1,225,000	76.42%
Born outside the United States	5,676,000	20.16%	4,535,000	21.60%	772,000	13.89%	369,000	23.02%
Missing	286,000	1.02%	132,000	0.63%	144,000	2.59%	9,000	0.56%

Source: 2015 Nonemployer data and 2015 Census Numident.

Note: This table presents firm-owner level place of birth information by firm type. Owners of sole proprietorships are identified from Form 1040. Owners of partnerships and S-corps are identified from Schedule K-1. Place of Birth information is obtained from the Census Numident. Only person owners are included in this table (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information). Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories. Missing indicates that an individual did not match to the Numident or they did match but there was no state of birth information available.

* Born in the United States include birthplace in Puerto Rico and U.S. Island Areas.

Table 37: Citizenship Status Distribution, Firm-owner Level, by Legal Form of Organization

	All Firm-Owner Pairs		Sole Proprietorships		Partnerships		S-Corps	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	28,160,000	100.00%	21,000,000	100.00%	5,557,000	100.00%	1,603,000	100.00%
U.S. Citizen	23,950,000	85.05%	17,680,000	84.19%	4,916,000	88.46%	1,356,000	84.59%
Non U.S. Citizen	3,527,000	12.52%	2,974,000	14.16%	357,000	6.42%	196,000	12.23%
Missing	682,000	2.42%	347,000	1.65%	284,000	5.11%	51,000	3.18%

Source: 2015 Nonemployer data and 2015 Census Numident.

Note: This table presents firm-owner level citizenship information by firm type. Owners of sole proprietorships are identified from Form 1040. Owners of partnerships and S-corps are identified from Schedule K-1. Citizenship information is obtained from the Census Numident. Only person owners are included in this table (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information). Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories. Missing indicates that an individual did not match to the Numident or they did match but we did not assign citizenship status. For individuals who are missing citizenship in the Numident, we assign the status "U.S. Citizen" unless they had a non-U.S. place of birth, in which case they are in the missing category (see text for more information).

Table 38: Place of Birth Distribution, Owner Level, by Legal Form of Organization

	Total		Born in the United States*		Born outside the United States		Missing	
	Number	Column Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All Groups	24,500,000	100.00%	19,220,000	78.45%	5,029,000	20.53%	251,000	1.02%
Sole Proprietorship (SP)	19,220,000	78.45%	14,930,000	77.68%	4,167,000	21.68%	128,000	0.67%
Partnership (P)	3,201,000	13.07%	2,646,000	82.66%	442,000	13.81%	113,000	3.53%
S-Corp (S)	1,007,000	4.11%	780,000	77.46%	220,000	21.85%	8,100	0.80%
SP and P	602,000	2.46%	518,000	86.05%	83,000	13.79%	850	0.14%
SP and S	265,000	1.08%	181,000	68.30%	83,000	31.32%	710	0.27%
P and S	159,000	0.65%	132,000	83.02%	27,000	16.98%	100	0.06%
SP, P, and S	41,500	0.17%	35,000	84.34%	6,100	14.70%	N < 15	(D)

Source: 2015 Nonemployer data and 2015 Census Numident.

Note: This table reports owner-level place of birth information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. Owners of sole proprietorships are identified from Form 1040. Owners of partnerships and S-corps are identified from Schedule K-1. Place of birth information is obtained from the Numident. The Missing indicates that an individual was not matched to the Numident or they matched but were missing state of birth information.

N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

* Born in the United States include birthplace in Puerto Rico and U.S. Island Areas

Table 39: Citizenship Distribution, Owner Level, by Legal Form of Organization

	Total		U.S. Citizen		Non U.S. Citizen		Missing	
	Number	Column Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All Groups	24,500,000	100.00%	20,750,000	84.69%	3,193,000	13.03%	557,000	2.27%
Sole Proprietorship (SP)	19,220,000	78.45%	16,150,000	84.03%	2,753,000	14.32%	318,000	1.65%
Partnership (P)	3,201,000	13.07%	2,803,000	87.57%	215,000	6.72%	183,000	5.72%
S-Corp (S)	1,007,000	4.11%	859,000	85.30%	116,000	11.52%	32,500	3.23%
SP and P	602,000	2.46%	548,000	91.03%	42,500	7.06%	11,500	1.91%
SP and S	265,000	1.08%	208,000	78.49%	52,000	19.62%	5,300	2.00%
P and S	159,000	0.65%	142,000	89.31%	11,000	6.92%	5,900	3.71%
SP, P, and S	41,500	0.17%	37,500	90.36%	2,700	6.51%	1,200	2.89%

Source: 2015 Nonemployer data and 2015 Census Numident.
Note: This table reports owner-level citizenship information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. Owners of sole proprietorships are identified from Form 1040. Owners of partnerships and S-corps are identified from Schedule K-1. Citizenship information is obtained from the Census Numident. Missing indicates that an individual did not match to the Numident or they did match but we did not assign citizenship status. For individuals who are missing citizenship in the Numident, we assign the status "U.S. Citizen" unless they had a non-U.S. place of birth, in which case they are in the missing category (see text for more information).

Table 40: Firm Ownership Place of Birth Distribution, by Legal Form of Organization & Method of Assignment

	Total	Born outside US		Born in US*		Equal		Unknown		Missing	
	Number (1)	Number (2a)	Row Percent (2b)	Number (3a)	Row Percent (3b)	Number (4a)	Row Percent (4b)	Number (5a)	Row Percent (5b)	Number (6a)	Row Percent (6b)
Partnerships and S-Corps											
Four Owners, 10% Rule (SBO)	2,787,000	475,000	17.04%	2,099,000	75.31%	88,000	3.16%	24,000	0.86%	101,500	3.64%
Four Owners, No 10% Rule	2,787,000	479,000	17.19%	2,140,000	76.79%	89,000	3.19%	24,500	0.88%	54,500	1.96%
All Owners, 10% Rule	2,787,000	478,000	17.15%	2,119,000	76.03%	87,500	3.14%	22,000	0.79%	80,000	2.87%
All Owners, No 10% Rule	2,787,000	482,000	17.29%	2,163,000	77.61%	88,000	3.16%	22,500	0.81%	30,500	1.09%
Partnerships											
Four Owners, 10% Rule (SBO)	1,681,000	208,000	12.37%	1,294,000	76.98%	68,500	4.07%	15,000	0.89%	96,000	5.71%
Four Owners, No 10% Rule	1,681,000	212,000	12.61%	1,334,000	79.36%	69,000	4.10%	15,500	0.92%	50,000	2.97%
All Owners, 10% Rule	1,681,000	211,000	12.55%	1,312,000	78.05%	67,500	4.02%	15,500	0.92%	75,500	4.49%
All Owners, No 10% Rule	1,681,000	215,000	12.79%	1,355,000	80.61%	68,000	4.05%	16,000	0.95%	27,000	1.61%
S-Corps											
Four Owners, 10% Rule (SBO)	1,106,000	267,000	24.14%	805,000	72.78%	20,000	1.81%	8,900	0.80%	5,700	0.52%
Four Owners, No 10% Rule	1,106,000	267,000	24.14%	806,000	72.88%	20,000	1.81%	8,900	0.80%	4,400	0.40%
All Owners, 10% Rule	1,106,000	268,000	24.23%	807,000	72.97%	20,000	1.81%	6,900	0.62%	4,900	0.44%
All Owners, No 10% Rule	1,106,000	268,000	24.23%	808,000	73.06%	20,000	1.81%	6,900	0.62%	3,600	0.33%

Source: 2015 Nonemployer data and 2015 Census Numident.

Note: This table reports firm level ownerships by place of birth for partnerships and S-corps. Owners are identified from Schedule K-1. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Firms are assigned place of birth using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. Depending on the method of assignment, a firm is assigned to a place of birth if members of that group account for more than 50 percent of ownership reported by either the four largest owners or all owners. Firms in which ownership is split 50-50 between U.S. born/Born outside U.S. are assigned to the "Equal" group. Firms in which no place of birth accounts for more than 50 percent of ownership are not assigned to a place of birth group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned to a place of birth category; where "No 10% Rule" is indicated, those firms are assigned to a place of birth category if possible. The "Unknown" category corresponds to those missing state of birth or have an edited place of birth. Ownership shares belonging to other firms do not figure into this process.

* Born in the United States include birthplace in Puerto Rico and U.S. Island Areas

Table 41: Firm Ownership Citizenship Distribution, by Legal Form of Organization & Method of Assignment

	Total	Not a U.S. Citizen		U.S. Citizen		Equal		Unknown		Missing	
	Number (1)	Number (2a)	Row Percent (2b)	Number (3a)	Row Percent (3b)	Number (4a)	Row Percent (4b)	Number (6a)	Row Percent (6b)	Number (6a)	Row Percent (6b)
Partnerships and S-Corps											
Four Owners, 10% Rule (SBO)	2,787,000	230,000	8.25%	2,322,000	83.32%	67,000	2.40%	48,500	1.74%	119,000	4.27%
Four Owners, No 10% Rule	2,787,000	232,000	8.32%	2,371,000	85.07%	67,500	2.42%	49,500	1.78%	67,000	2.40%
All Owners, 10% Rule	2,787,000	230,000	8.25%	2,326,000	83.46%	67,000	2.40%	48,000	1.72%	116,000	4.16%
All Owners, No 10% Rule	2,787,000	232,000	8.32%	2,372,000	85.11%	67,000	2.40%	49,000	1.76%	66,500	2.39%
Partnerships											
Four Owners, 10% Rule (SBO)	1,681,000	87,500	5.21%	1,413,000	84.06%	50,000	2.97%	24,500	1.46%	106,000	6.31%
Four Owners, No 10% Rule	1,681,000	89,000	5.29%	1,461,000	86.91%	50,000	2.97%	25,500	1.52%	55,000	3.27%
All Owners, 10% Rule	1,681,000	87,500	5.21%	1,417,000	84.30%	49,500	2.94%	24,000	1.43%	103,000	6.13%
All Owners, No 10% Rule	1,681,000	89,000	5.29%	1,463,000	87.03%	50,000	2.97%	24,500	1.46%	54,500	3.24%
S-Corps											
Four Owners, 10% Rule (SBO)	1,106,000	143,000	12.93%	909,000	82.19%	17,000	1.54%	24,500	2.22%	13,000	1.18%
Four Owners, No 10% Rule	1,106,000	143,000	12.93%	910,000	82.28%	17,500	1.58%	24,500	2.22%	12,000	1.08%
All Owners, 10% Rule	1,106,000	143,000	12.93%	909,000	82.19%	17,000	1.54%	24,500	2.22%	13,000	1.18%
All Owners, No 10% Rule	1,106,000	143,000	12.93%	910,000	82.28%	17,000	1.54%	24,500	2.22%	12,000	1.08%

Source: 2015 Nonemployer data and 2015 Census Numident.

Note: This table reports firm level ownerships by citizenship status for partnerships and S-corps. Owners are identified from Schedule K-1. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Firms are assigned citizenship status using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. Depending on the method of assignment, a firm is assigned to a citzenship status if members of that group account for more than 50 percent of ownership reported by either the four largest owners or all owners. Firms in which ownership is split 50-50 between U.S. citizens/non citizens are assigned to the "Equal" group. Firms in which no citizenship status accounts for more than 50 percent of ownership are not assigned to a citizenship group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned to a citizenship category; where "No 10% Rule" is indicated, those firms are assigned to a citizenship category if possible. Ownership shares belonging to other firms do not figure into this process.

Table 42: Age Distribution, Firm-Owner Pairs, by Legal Form of Organization

	All Firm-Owner Pairs		Sole Proprietorships		Partnerships		S-Corps	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	28,160,000	100%	21,000,000	100%	5,557,000	100%	1,603,000	100%
Under 30	3,305,000	11.74%	2,974,000	14.16%	272,000	4.89%	59,500	3.71%
30-54	14,410,000	51.17%	11,270,000	53.67%	2,348,000	42.25%	794,000	49.53%
55+	10,160,000	36.08%	6,631,000	31.58%	2,793,000	50.26%	741,000	46.23%
Missing	280,000	0.99%	127,000	0.60%	144,000	2.59%	8,600	0.54%

Source: 2015 Nonemployer database, 2014-2016 K-1 data, and Census Numident.

Note: This table presents firm-owner level age information by firm type. Owners of sole proprietorships are identified from Form 1040 (via the Business Register nonemployer extract). Owners of partnerships and s-corps are identified from Schedule K-1. Age is calculated for 2015 from date of birth information obtained from the Numident. Only person owners are included in this table (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information). Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories.

Table 43: Age Distribution, Owner Level, by Legal Forms of Organization

	Total	Under 30		30-54		55+		Missing	
	Number	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All Groups	24,500,000	3,118,000	12.73%	12,680,000	51.76%	8,455,000	34.51%	245,000	1.00%
Sole Proprietorship (SP)	19,220,000	2,832,000	14.73%	10,290,000	53.54%	5,975,000	31.09%	123,000	0.64%
Partnership (P)	3,201,000	199,000	6.22%	1,361,000	42.52%	1,528,000	47.74%	113,000	3.53%
S-Corp (S)	1,007,000	39,500	3.92%	495,000	49.16%	465,000	46.18%	7,900	0.78%
SP and P	602,000	29,500	4.90%	294,000	48.84%	278,000	46.18%	500	0.08%
SP and S	265,000	13,500	5.09%	154,000	58.11%	97,000	36.60%	500	0.19%
P and S	159,000	3,200	2.01%	67,500	42.45%	88,500	55.66%	100	0.06%
SP, P, and S	41,500	700	1.69%	18,000	43.37%	22,500	54.22%	<15	(D)

Source: 2015 Nonemployer database, 2014-2016 K-1 data, and Census Numident.

Note: This table reports owner-level age information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. Owners of sole proprietorships are identified from Form 1040 (via the Business Register nonemployer extract). Owners of partnerships and s-corps are identified from Schedule K-1. Age is calculated for 2015 from date of birth information obtained from the Numident.

N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table 44: Firm Ownership Age Distribution, by Legal Form of Organization and Method of Assignment

	Total	Under 30		30-54		55+		Missing	
	Number	Number	Row Percent	Number	Row Percent	Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All									
Four Owners, 10% Rule (SBO)	2,787,000	67,000	2.40%	1,159,000	41.59%	1,207,000	43.31%	354,000	12.70%
Four Owners, No 10% Rule	2,787,000	68,500	2.46%	1,183,000	42.45%	1,244,000	44.64%	290,000	10.41%
All Owners, 10% Rule	2,787,000	69,000	2.48%	1,172,000	42.05%	1,212,000	43.49%	333,000	11.95%
All Owners, No 10% Rule	2,787,000	70,000	2.51%	1,186,000	42.55%	1,244,000	44.64%	286,000	10.26%
Partnerships									
Four Owners, 10% Rule (SBO)	1,681,000	34,000	2.02%	616,000	36.64%	729,000	43.37%	301,000	17.91%
Four Owners, No 10% Rule	1,681,000	36,500	2.17%	640,000	38.07%	766,000	45.57%	238,000	14.16%
All Owners, 10% Rule	1,681,000	36,000	2.14%	629,000	37.42%	735,000	43.72%	281,000	16.72%
All Owners, No 10% Rule	1,681,000	37,000	2.20%	643,000	38.25%	766,000	45.57%	235,000	13.98%
S-Corps									
Four Owners, 10% Rule (SBO)	1,106,000	33,000	2.98%	543,000	49.10%	478,000	43.22%	53,000	4.79%
Four Owners, No 10% Rule	1,106,000	33,000	2.98%	543,000	49.10%	479,000	43.31%	52,000	4.70%
All Owners, 10% Rule	1,106,000	33,000	2.98%	543,000	49.10%	478,000	43.22%	53,000	4.79%
All Owners, No 10% Rule	1,106,000	33,000	2.98%	543,000	49.10%	479,000	43.31%	51,500	4.66%

Source: 2015 Nonemployer database, 2014-2016 K-1 data, and Census Numident.

Note: This table reports firm level ownerships by age group for partnerships and s-corps. Owners are identified from Schedule K-1. Firms not matched to K-1s or matched to zero person owners are excluded from this analysis. Firms are assigned to age groups using the indicated number of person owners with the largest individual ownership shares, which are obtained from Schedule K-1. A firm is assigned to an age group if members of that group account for more than 50 percent of ownership reported by the indicated number of owners. Firms in which no age group accounts for more than 50 percent of ownership are not assigned to an age group. Where "10% Rule" is indicated, firms in which the person owner with the largest ownership share owns less than ten percent of the firm are not assigned to an age category; where "No 10% Rule" is indicated, those firms are assigned to an age category if possible. Ownership shares belonging to other firms do not figure into this process. Age for 2015 is calculated using date of birth information obtained from the Numident.

Appendix Tables

Table A. 1: Availability of AR Composite Race & Hispanic Origin Data, Firm-owner level, by LFO

2015 Nonemployers Database	Firm-PIK Pairs	Owner Match to AR Composite File		Availability of Race Response not Including Some Other Race		Availability of Some Other Race Response		Availability of Hispanic Origin Response	
	Number	Number	As % of Firm-PIK Pairs	Number	As % of Firm-PIK Pairs	Number	As % of Firm-PIK Pairs	Number	As % of Firm-PIK Pairs
	(1)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Total	28,160,000	27,370,000	97.19	25,765,000	91.50	899,000	3.19	27,120,000	96.31
Sole Proprietorships	21,000,000	20,404,000	97.16	18,900,000	90.00	816,000	3.89	20,210,000	96.24
Partnerships	5,557,000	5,382,000	96.85	5,285,000	95.11	54,000	0.97	5,339,000	96.08
S-corps	1,603,000	1,581,000	98.63	1,583,000	98.75	29,500	1.84	1,567,000	97.75

Source: 2015 Nonemployer data and Administrative Records composite file.

Table A. 2: Race Distribution, Firm-owner level, by Legal Form of Organization

	2015 PCR Data		AR Composite File		Difference (AR – PCR)	
	Number	Column %	Number	Column %	Number	Column %
Total	28,160,000	100.00	28,160,000	100.00		
White	20,890,000	74.18	21,020,000	74.65	130,000	0.47
Black	2,363,000	8.39	2,523,000	8.96	160,000	0.57
American Indian Alaska Native	50,000	0.18	178,000	0.63	128,000	0.45
Asian	1,531,000	5.44	1,516,000	5.38	-15,000	-0.06
Native Hawaiian or Pacific Islander	24,500	0.09	33,500	0.12	9,000	0.03
Multiple Races	456,000	1.62	439,000	1.56	-17,000	-0.06
Missing	2,847,000	10.11	2,451,000	8.70	-396,000	-1.41
Sole Proprietorships						
Total	21,000,000	100.00	21,000,000	100.00		
White	14,840,000	70.66	14,910,000	70.99	70,000	0.33
Black	2,221,000	10.58	2,374,000	11.30	153,000	0.72
American Indian Alaska Native	44,500	0.21	156,000	0.74	111,500	0.53
Asian	1,073,000	5.11	1,062,000	5.06	-11,000	-0.05
Native Hawaiian or Pacific Islander	21,000	0.10	29,000	0.14	8,000	0.04
Multiple Races	381,000	1.81	366,000	1.74	-15,000	-0.07
Missing	2,421,000	11.53	2,105,000	10.02	-316,000	-1.51
Partnerships & S-corps						
Total	7,160,000	100.00	7,160,000	100.00		
White	6,050,000	84.50	6,111,000	85.36	61,000	0.86
Black	142,000	1.98	149,000	2.08	7,000	0.10
American Indian Alaska Native	5,400	0.07	22,000	0.31	16,600	0.24
Asian	458,000	6.40	454,000	6.34	-4,000	-0.06
Native Hawaiian or Pacific Islander	3,200	0.05	4,300	0.06	1,100	0.01
Multiple Races	75,500	1.06	73,500	1.02	-2,000	-0.04
Missing	426,000	5.95	346,000	4.84	-80,000	-1.11
Partnerships						
Total	5,557,000	100.00	5,557,000	100.00		
White	4,738,000	85.26	4,785,000	86.10	47,000	0.84
Black	84,500	1.52	89,000	1.60	4,500	0.08
American Indian Alaska Native	3,700	0.07	15,500	0.28	11,800	0.21
Asian	345,000	6.20	341,000	6.14	-4,000	-0.06
Native Hawaiian or Pacific Islander	2,300	0.04	3,000	0.05	700	0.01
Multiple Races	53,500	0.96	52,000	0.93	-1,500	-0.03
Missing	330,000	5.95	272,000	4.90	-58,000	-1.05
S-corps						
Total	1,603,000	100.00	1,603,000	100.00		
White	1,312,000	81.84	1,327,000	82.79	15,000	0.95
Black	57,500	3.58	60,000	3.74	2,500	0.16
American Indian Alaska Native	1,600	0.10	6,600	0.41	5,000	0.31
Asian	114,000	7.10	113,000	7.02	-1,000	-0.08
Native Hawaiian or Pacific Islander	900	0.06	1,300	0.08	400	0.02
Multiple Races	22,000	1.38	21,500	1.34	-500	-0.04
Missing	95,500	5.95	74,000	4.62	-21,500	-1.33

Source: 2015 Nonemployer data, 2015 Previous Census Records (PCR) file, Administrative Records Composite file.

* Some Other Race is not included in the distribution.

Table A.3: Race Distribution, Firm-owner Level, by LFO (excluding missing category)

	2015 PCR Data		AR Composite File		Difference (AR - PCR)	
	Number	Column %	Number	Column %	Number	Column %
Total	25,320,000	100.00	25,710,000	100.00		
White	20,890,000	82.50	21,020,000	81.76	130,000	-0.74
Black	2,363,000	9.33	2,523,000	9.81	160,000	0.48
American Indian Alaska Native	50,000	0.20	178,000	0.69	128,000	0.49
Asian	1,531,000	6.05	1,516,000	5.90	-15,000	-0.15
Native Hawaiian or Pacific Islander	24,500	0.10	33,500	0.13	9,000	0.03
Multiple Races	456,000	1.80	439,000	1.71	-17,000	-0.09
Sole Proprietorships						
Total	18,580,000	100.00	18,900,000	100.00		
White	14,840,000	79.87	14,910,000	78.90	70,000	-0.97
Black	2,221,000	11.95	2,374,000	12.56	153,000	0.61
American Indian Alaska Native	44,500	0.24	156,000	0.82	111,500	0.58
Asian	1,073,000	5.78	1,062,000	5.62	-11,000	0.16
Native Hawaiian or Pacific Islander	21,000	0.11	29,000	0.15	8,000	0.04
Multiple Races	381,000	2.05	366,000	1.94	-15,000	-0.11
Partnerships & S-corps						
Total	6,734,000	100.00	6,813,000	100.00		
White	6,050,000	89.84	6,111,000	89.70	61,000	-0.14
Black	142,000	2.11	149,000	2.19	7,000	0.08
American Indian Alaska Native	5,400	0.08	22,000	0.32	16,600	0.24
Asian	458,000	6.80	454,000	6.66	-4,000	-0.14
Native Hawaiian or Pacific Islander	3,200	0.05	4,300	0.06	1,100	0.01
Multiple Races	75,500	1.12	73,500	1.08	-2,000	-0.04
Partnerships						
Total	5,227,000	100.00	5,285,000	100.00		
White	4,738,000	90.64	4,785,000	90.54	47,000	-0.10
Black	84,500	1.62	89,000	1.68	4,500	0.06
American Indian Alaska Native	3,700	0.07	15,500	0.29	11,800	0.22
Asian	345,000	6.60	341,000	6.45	-4,000	-0.15
Native Hawaiian or Pacific Islander	2,300	0.04	3,000	0.06	700	0.02
Multiple Races	53,500	1.02	52,000	0.98	-1,500	-0.04
S-corps						
Total	1,507,000	100.00	1,529,000	100.00		
White	1,312,000	87.06	1,327,000	86.80	15,000	-0.26
Black	57,500	3.82	60,000	3.92	2,500	0.10
American Indian Alaska Native	1,600	0.11	6,600	0.43	5,000	0.32
Asian	114,000	7.56	113,000	7.36	-1,000	-0.20
Native Hawaiian or Pacific Islander	900	0.06	1,300	0.08	400	0.02
Multiple Races	22,000	1.46	21,500	1.41	-500	-0.05

Table A.4: Hispanic Origin Distribution, Firm-owner level, by LFO

	2015 PCR Data		AR Composite File		Difference (AR - PCR)	
	Number	Column %	Number	Column %	Number	Column %
Hispanic Origin						
Total	28,160,000	100.00	28,160,000	100.00		
Hispanic	2,920,000	10.37	3,844,000	13.65	924,000	3.28
Nonhispanic	23,410,000	83.13	23,270,000	82.63	-140,000	-0.50
Missing	1,829,000	6.50	1,045,000	3.71	-784,000	-2.79
Sole Proprietorships						
Total	21,000,000	100.00	21,000,000	100.00		
Hispanic	2,595,000	12.36	3,356,000	15.98	761,000	3.62
Nonhispanic	16,930,000	80.62	16,860,000	80.29	-70,000	-0.33
Missing	1,481,000	7.05	792,000	3.77	-689,000	-3.28
Partnerships & S-corps						
Total	7,160,000	100.00	7,160,000	100.00		
Hispanic	325,000	4.54	488,000	6.82	163,000	2.28
Nonhispanic	6,487,000	90.60	6,418,000	89.64	-69,000	-0.96
Missing	349,000	4.87	254,000	3.55	-95,000	-1.32
Partnerships						
Total	5,557,000	100.00	5,557,000	100.00		
Hispanic	193,000	3.47	310,000	5.58	117,000	2.11
Nonhispanic	5,081,000	91.43	5,029,000	90.50	-52,000	-0.93
Missing	282,000	5.07	218,000	3.92	-64,000	-1.15
S-corps						
Total	1,603,000	100.00	1,603,000	100.00		
Hispanic	131,000	8.17	178,000	11.10	47,000	2.93
Nonhispanic	1,405,000	87.65	1,389,000	86.65	-16,000	-1.00
Missing	66,000	4.12	35,500	2.21	-30,500	-1.91

Source: 2015 Nonemployer data, 2015 Previous Census Records (PCR) file, Administrative Records Composite file

Table A.5: Hispanic Origin Distribution, Firm-owner Level, by LFO (excludes missing category)

	2015 PCR Data		AR Composite File		Difference (AR - PCR)	
	Number	Column %	Number	Column %	Number	Column %
Hispanic Origin						
Total	26,330,000	100.00	27,120,000	100.00		
Hispanic	2,920,000	11.09	3,844,000	14.17	924,000	3.08
Non-Hispanic	23,410,000	88.91	23,270,000	85.80	-140,000	-3.11
Sole Proprietorships						
Total	19,520,000	100.00	20,210,000	100.00		
Hispanic	2,595,000	13.29	3,356,000	16.61	761,000	3.32
Non-Hispanic	16,930,000	89.73	16,860,000	83.42	-70,000	-6.31
Partnerships & S-corps						
Total	6,811,000	100.00	6,906,000	100.00		
Hispanic	325,000	4.77	488,000	7.07	163,000	2.30
Non-Hispanic	6,487,000	95.24	6,418,000	92.93	-69,000	-2.31
Partnerships						
Total	5,275,000	100.00	5,339,000	100.00		
Hispanic	193,000	3.66	310,000	5.81	117,000	2.15
Non-Hispanic	5,081,000	96.32	5,029,000	94.19	-52,000	-2.13
S-corps						
Total	1,536,000	100.00	1,567,000	100.00		
Hispanic	131,000	8.53	178,000	11.36	47,000	2.83
Non-Hispanic	1,405,000	91.47	1,389,000	88.64	-16,000	-2.83

Source: 2015 Nonemployer data, 2015 Previous Census Records (PCR) file, Best Race Administrative Records file

Table A.6: Availability of AR Composite File Race & Hispanic Origin Data, Owner Level, by LFO

2015 Nonemployers Database	Owners with Unique PIKs	Owner Match to AR File		Availability of Race Response not Including Some Other Race		Availability of Some Other Race Response		Availability of Hispanic Origin Response	
	Number	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs	Number	As % of firm-PIK pairs
Total	24,500,000	23,770,000	97.42	22,220,000	91.07	844,000	3.46	23,550,000	96.52
Partnership	3,201,000	3,067,000	95.81	3,004,000	93.85	35,000	1.09	3,041,000	95.00
Partnership and S-corp	159,000	159,000	100.00	156,000	98.11	1,500	0.94	158,000	99.37
Sole proprietorship	19,220,000	18,660,000	97.09	17,230,000	89.65	774,000	4.03	18,480,000	96.15
Sole proprietorship and Partnership	602,000	598,000	99.34	585,000	97.18	6,800	1.13	594,000	98.67
Sole proprietorship, Partnership, and S-corp	41,500	41,000	98.80	40,500	97.59	350	0.84	41,000	98.80
Sole proprietorship and S-corp	265,000	259,000	97.74	246,000	92.83	6,600	2.49	256,000	96.60
S-corp	1,007,000	994,000	98.71	960,000	95.33	19,500	1.94	985,000	97.82

Source: 2015 Nonemployer data and Best Race Administrative Records file.
Note: This table reports owner-level race information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category.

Table A.7: Race Distribution, Owner Level, by LFO - AR Composite File

2015 Nonemployers Database	Total	White		Black		American Indian or Alaska Native		Asian		Native Hawaiian or Pacific Islander		Multiple Races		Missing	
	Number	Number	Pct	Number	Pct	Number	Pct	Number	Pct	Number	Pct	Number	Pct	Number	Pct
Total	24,500,000	17,941,000	73.23	2,392,000	9.76	164,800	0.67	1,297,400	5.30	31,190	0.13	396,600	1.62	2,273,100	9.28
Partnership	3,201,000	2,695,000	84.19	61,000	1.91	10,500	0.33	205,000	6.40	2,100	0.07	31,500	0.98	196,000	6.12
Partnership and S-Corp	159,000	141,000	88.68	1,900	1.19	400	0.25	11,000	6.92	70	0.04	1,500	0.94	3,100	1.95
Sole proprietorship	19,220,000	13,500,000	70.24	2,260,000	11.76	146,000	0.76	957,000	4.98	27,500	0.14	339,000	1.76	1,991,000	10.36
Sole propiertorship and Partnership	602,000	529,000	87.87	14,000	2.33	2,100	0.35	34,000	5.65	400	0.07	6,500	1.08	16,000	2.66
Sole proprietorship, Partnership, and S-Corp	41,500	37,000	89.16	600	1.45	100	0.24	2,400	5.78	20	0.05	400	0.96	1,000	2.41
Sole proprietorship and S-Corp	265,000	210,000	79.25	12,500	4.72	1,200	0.45	18,000	6.79	250	0.09	4,200	1.58	19,000	7.17
S-Corp	1,007,000	829,000	82.32	42,000	4.17	4,500	0.45	70,000	6.95	850	0.08	13,500	1.34	47,000	4.67

Source: 2015 Nonemployer data and Administrative Records Composite file.

Note: This table reports owner-level race information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. The “Missing” category indicates that an individual was either not matched to the AR Composite file or the race information was missing.

Table A.8: Hispanic Origin Distribution, Owner Level, by LFO - AR Composite File

2015 Nonemployers Database	Total	Hispanic		Non-Hispanic		Missing	
	Number	Number	Percent	Number	Percent	Number	Percent
Total	24,500,000	3,549,400	14.49	20,003,500	81.65	947,100	3.87
Partnership	3,201,000	194,000	6.06	2,847,000	88.94	160,000	5.00
Partnership and S-corp	159,000	9,500	5.97	148,000	93.08	1,500	0.94
Sole proprietorship	19,220,000	3,149,000	16.38	15,330,000	79.76	741,000	3.86
Sole propiertorship and Partnership	602,000	40,000	6.64	554,000	92.03	8,000	1.33
Sole proprietorship, Partnership, and S-corp	41,500	2,400	5.78	38,500	92.77	600	1.45
Sole proprietorship and S-corp	265,000	41,500	15.66	214,000	80.75	9,500	3.58
S-corp	1,007,000	113,000	11.22	872,000	86.59	22,000	2.18

Source: 2015 Nonemployer data and Administrative Records Composite file.
Note: This table reports owner-level ethnicity information by type(s) of nonemployer firm owned. Each individual owner appears exactly once in exactly one ownership category. The “Missing” category indicates that an individual was either not matched to the AR Composite file or the Hispanic origin information was missing..

Table A. 9: Firm Ownership Race Distribution, by LFO & Method of Assignment - AR Composite File

	Total	White		Black		AIAN		Asian		NHPI		Multiple Races		Not Assigned		Unlinkable	
	N	N	Row %	N	Row %	N	Row %	N	Row %	NHPI	Row %	N	Row %	N	Row %	N	Row %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
All																	
Four Owners, 10% Rule (SBO)	23,790,000	17,180,000	72.22%	2,446,000	10.28%	162,000	0.68%	1,222,000	5.14%	30,500	0.13%	387,000	1.63%	1,731,000	7.28%	629,000	2.64%
Four Owners, No 10% Rule	23,790,000	17,230,000	72.43%	2,447,000	10.29%	162,000	0.68%	1,224,000	5.15%	30,500	0.13%	388,000	1.63%	1,680,000	7.06%	629,000	2.64%
All Owners, 10% Rule	23,790,000	17,190,000	72.26%	2,446,000	10.28%	162,000	0.68%	1,222,000	5.14%	30,500	0.13%	387,000	1.63%	1,728,000	7.26%	629,000	2.64%
All Owners, No 10% Rule	23,790,000	17,230,000	72.43%	2,447,000	10.29%	162,000	0.68%	1,223,000	5.14%	30,500	0.13%	388,000	1.63%	1,680,000	7.06%	629,000	2.64%
Sole Proprietorships	21,000,000	14,910,000	71.00%	2,374,000	11.30%	156,000	0.74%	1,062,000	5.06%	29,000	0.14%	366,000	1.74%	1,506,000	7.17%	599,000	2.85%
Partnerships																	
Four Owners, 10% Rule (SBO)	1,681,000	1,378,000	81.98%	26,500	1.58%	2,500	0.15%	85,500	5.09%	450	0.03%	8,100	48.00%	165,000	9.82%	15,500	0.92%
Four Owners, No 10% Rule	1,681,000	1,425,000	84.77%	26,500	1.58%	2,600	0.15%	87,000	5.18%	450	0.03%	8,300	49.00%	115,000	6.84%	15,500	0.92%
All Owners, 10% Rule	1,681,000	1,381,000	82.15%	26,500	1.58%	2,500	0.15%	85,500	5.09%	450	0.03%	8,100	48.00%	161,000	9.58%	15,000	0.89%
All Owners, No 10% Rule	1,681,000	1,426,000	84.83%	26,500	1.58%	2,500	0.15%	87,000	5.18%	450	0.03%	8,200	49.00%	114,000	6.78%	15,000	0.89%
S-Corps																	
Four Owners, 10% Rule (SBO)	1,106,000	893,000	80.74%	45,500	4.11%	3,900	0.35%	74,000	6.69%	750	0.07%	13,500	1.22%	60,500	5.47%	14,500	1.31%
Four Owners, No 10% Rule	1,106,000	894,000	80.83%	45,500	4.11%	3,900	0.35%	74,000	6.69%	750	0.07%	13,500	1.22%	59,500	5.38%	14,500	1.31%
All Owners, 10% Rule	1,106,000	893,000	80.74%	45,500	4.11%	3,900	0.35%	74,000	6.69%	750	0.07%	13,500	1.22%	60,500	5.47%	14,500	1.31%
All Owners, No 10% Rule	1,106,000	894,000	80.83%	45,500	4.11%	3,900	0.35%	74,000	6.69%	750	0.07%	13,500	1.22%	59,000	5.33%	14,500	1.31%

Source: 2015 Nonemployer data, Administrative Records Composite file.

Table A. 10: Firm Ownership Hispanic Origin Distribution, by LFO & Method of Assignment - AR Composite File

	Total	Hispanic		Non-Hispanic		Equal		Not Assigned		Unlinkable	
	Number	Number	Row %	Number	Row %	Number	Row %	Number	Row %	Number	Row %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
All											
Four Owners, 10% Rule (SBO)	23,790,000	3,564,000	14.98%	19,230,000	80.83%	51,500	0.22%	311,000	1.31%	629,000	2.64%
Four Owners, No 10% Rule	23,790,000	3,565,000	14.99%	19,280,000	81.04%	52,000	0.22%	260,000	1.09%	629,000	2.64%
All Owners, 10% Rule	23,790,000	3,564,000	14.98%	19,240,000	80.87%	51,000	0.21%	308,000	1.29%	629,000	2.64%
All Owners, No 10% Rule	23,790,000	3,565,000	14.99%	19,290,000	81.08%	51,500	0.22%	259,000	1.09%	629,000	2.64%
Sole Proprietorships	21,000,000	3,356,000	15.98%	16,860,000	80.29%	0	0.00%	193,000	0.92%	599,000	2.85%
Partnerships											
Four Owners, 10% Rule (SBO)	1,681,000	80,500	4.79%	1,445,000	85.96%	40,000	2.38%	10,000	0.59%	15,500	0.92%
Four Owners, No 10% Rule	1,681,000	82,000	4.88%	1,493,000	88.82%	40,500	2.41%	49,500	2.94%	15,500	0.92%
All Owners, 10% Rule	1,681,000	80,500	4.79%	1,449,000	86.20%	39,500	2.35%	96,500	5.74%	15,000	0.89%
All Owners, No 10% Rule	1,681,000	82,000	4.88%	1,494,000	88.88%	40,000	2.38%	49,000	2.91%	15,000	0.89%
S-corps											
Four Owners, 10% Rule (SBO)	1,106,000	128,000	11.57%	934,000	84.45%	11,500	1.04%	18,500	1.67%	14,500	1.31%
Four Owners, No 10% Rule	1,106,000	128,000	11.57%	935,000	84.54%	11,500	1.04%	17,000	1.54%	14,500	1.31%
All Owners, 10% Rule	1,106,000	128,000	11.57%	934,000	84.45%	11,500	1.04%	18,500	1.67%	14,500	1.31%
All Owners, No 10% Rule	1,106,000	128,000	11.57%	935,000	84.54%	11,500	1.04%	17,000	1.54%	14,500	1.31%

Source: 2015 Nonemployer data, Administrative Records Composite data

Table A. 11: Firm Ownership Race Distribution in the SBO, Previous Census Records (PCR) File & AR Composite File

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO		Total - 2015 Previous Census Records, 10% Rule		Total - 2015 Previous Census Records, No 10% Rule		Total - 2015 AR Composite, 10% Rule		Total - 2015 AR Composite, No 10% Rule	
	Number	Col %	Number	Col %	Number	Col %	Number	Col %	Number	Col %	Number	Col %	Number	Col %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(7)	(8)	(9)	(10)
Total	17,277,031	100.0	21,236,766	100.0	22,405,695	100.0	21,080,000	100.0	21,128,000	100.0	21,427,500	100.00	21,480,000	100.00
White	15,187,720	87.91	17,955,403	84.55	17,101,796	76.33	17,090,000	81.07	17,130,000	81.08	17,180,000	80.18	17,230,000	80.21
Black	1,103,049	6.38	1,815,298	8.55	2,475,266	11.05	2,290,000	10.86	2,291,000	10.84	2,446,000	11.42	2,447,000	11.39
American Indian or Alaska Native	176,889	1.02	213,029	1.00	246,740	1.10	46,000	0.22	46,000	0.22	162,000	0.76	162,000	0.75
Asian	784,118	4.54	1,152,134	5.43	1,436,876	6.41	1,234,000	5.85	1,236,000	5.85	1,222,000	5.70	1,224,000	5.70
Native Hawaiian and Pacific Islander	25,255	0.15	33,536	0.16	50,043	0.22	22,000	0.10	22,000	0.10	30,500	0.14	30,500	0.14
Some Other Race	N/A	N/A	67,366	0.32	1,094,974	4.89	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Multiple Races	N/A	N/A	N/A	N/A	N/A	N/A	403,000	1.91	403,000	1.91	387,000	1.81	388,000	1.81

Source: 2002 SBO, 2007 SBO, 2012 SBO, 2015 Previous Census Records file, 2015 Administrative Records Composite file.
Note: SBO percentages may sum greater than 100% because SBO firms can be assigned more than one race. 2015 figures use the four owners rules.
N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

A. 12: Firm Ownership by Hispanic Origin in the SBO, Previous Census Records (PCR) File & AR Composite File

	Total - 2002 SBO		Total - 2007 SBO		Total - 2012 SBO		Total - 2015 Previous Census Records, 10% Rule		Total - 2015 Previous Census Records, No 10% Rule		Total - 2015 AR Composite, 10% Rule		Total - 2015 AR Composite, No 10% Rule	
	Number	Col %	Number	Col %	Number	Col %	Number	Col %	Number	Col %	Number	Col %	Number	Col %
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(7)	(8)	(9)	(10)
Total	17,207,068	100.0	20,908,714	100.0	21,951,962	100.0	22,080,000	100.0	22,130,000	100.0	22,800,000	100.0	22,850,000	100.0
Hispanic	1,373,922	7.98	2,011,417	9.62	3,018,371	13.75	2,743,000	12.42	2,744,000	12.40	3,564,000	15.60	3,565,000	15.60
Non-Hispanic	15,833,146	92.02	18,897,297	90.38	18,933,591	86.25	19,330,000	87.55	19,380,000	87.57	19,230,000	84.30	19,280,000	84.40
Equal	0	0.00	196,178	0.94	91,214	0.42	34,000	0.15	34,500	0.16	51,500	0.20	52,000	0.20

Source: 2002 SBO, 2007 SBO, 2012 SBO, 2015 Previous Census Records file, 2015 Administrative Records Composite file.
Note: SBO percentages may sum greater than 100% because SBO firms can be assigned more than one race. 2015 figures use the four owners rules.

Table A. 13: Veteran Status Distribution, Owner Level, in the SBO and the 2015 VA USVETS Data

2012 SBO Veteran Status	2015 VA USVETS Veteran Status				
	Total	Veteran		Non-veteran	
		Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)
Total	814,000	81,500	10.0%	732,000	89.9%
Veteran	89,500	39,000	43.6%	50,500	56.4%
Non-veteran	724,000	42,500	5.9%	682,000	94.2%

Source: 2012 SBO and 2015 VA USVETS.

Note: This table reports unique owner-level comparison of veteran status between 2012 SBO and 2015 VA USVETS. Only observations from 2012 SBO are included. The sample consists of reported cases of in-scope firms for the owner of non-employer sole proprietors. As SBO includes PIKs for only soleproprietors, we are not able to examine other owners' veteran status responses.

Table A. 14: Veteran Status Distribution, Owner Level, in the 2015 ACS and the 2015 VA USVETS Data

2015 ACS Veteran Status	2015 VA USVETS Veteran Status				
	Total	Veteran		Non-veteran	
		Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)
Total	4,966,000	339,000	6.8%	4,627,000	93.2%
Now on active duty	14,000	7,000	50.0%	6,900	49.3%
On active duty in the past, but not now	347,000	263,000	75.8%	84,000	24.2%
Only on active duty for training in the Reserves or National Guard	61,000	24,000	39.3%	37,000	60.7%
Never served in the military	3,569,000	45,000	1.3%	3,524,000	98.7%
N/A	975,000	450	0.1%	974,000	99.9%

Source: 2015 ACS and 2015 VA USVETS.

Note: This table reports unique person-level comparison of veteran status between 2015 ACS and 2015 VA USVETS. Only observations from 2015 ACS are included. ACS asks about military service everyone who is 17 years old or older.
N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table A. 15: Veteran Status Distribution, Owner Level, in the 2015 ACS and the 2015 VA USVETS Data (reported cases only)

2015 ACS Veteran Status	2015 VA USVETS Veteran Status				
	Total	Veteran		Non-veteran	
		Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)
Total	4,678,000	314,000	6.7%	4,365,000	93.3%
Now on active duty	13,500	6,900	51.1%	6,500	48.2%
On active duty in the past, but not now	317,000	252,000	79.5%	65,000	20.5%
Only on active duty for training in the Reserves or National Guard	57,000	23,500	41.2%	33,500	58.8%
Never served in the military	3,316,000	30,500	0.9%	3,285,000	99.1%
N/A	975,000	450	0.1%	974,000	99.9%

Source: 2015 ACS and 2015 VA USVETS.

Note: This table reports unique person-level comparison of veteran status between 2015 ACS and 2015 VA USVETS. Only reported observations from 2015 ACS are included. Imputations are excluded.
N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table A. 16: Veteran StatusDistribution, Owner Level, in the 2015 ACS and the 2015 VA USVETS Data (reported cases only, excluding proxy responses)

2015 ACS Veteran Status	2015 VA USVETS Veteran Status				
	Total	Veteran		Non-veteran	
		Number	Row Percent	Number	Row Percent
	(1)	(2)	(3)	(4)	(5)
Total	1,903,000	205,000	10.8%	1,698,000	89.2%
Now on active duty	7,100	4,400	62.0%	2,600	36.6%
On active duty in the past, but not now	217,000	171,000	78.8%	46,000	21.2%
Only on active duty for training in the Reserves or National Guard	37,500	13,500	36.0%	24,000	64.0%
Never served in the military	1,641,000	16,000	1.0%	1,625,000	99.0%
N/A	200	<15	(D)	(D)	(D)

Source: 2015 ACS and 2015 VA USVETS.
Note: This table reports unique person-level comparison of veteran status between 2015 ACS and 2015 VA USVETS. Only reported observations from 2015 ACS are included. Imputations and proxy responses are excluded.
N/A: Not applicable; D: Withheld to avoid disclosing data for individual companies; data are included in higher level totals.

Table A. 17: Citizenship Status Distribution, Owner Level, by LFO

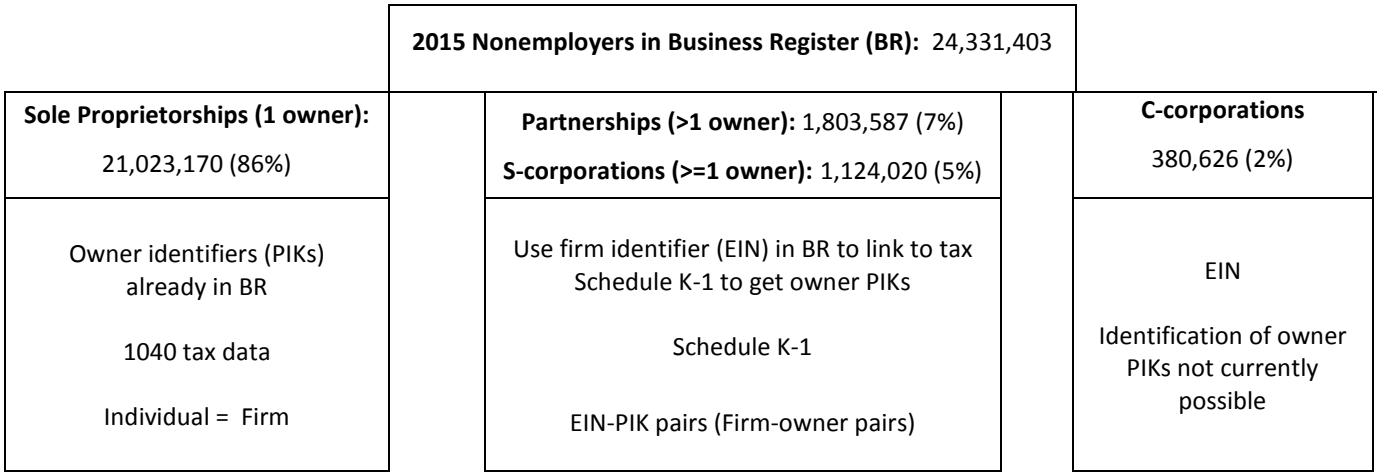
	All Firm-Owner Pairs		Sole Proprietorships		Partnerships		S-Corps	
	Number	Column Percent	Number	Column Percent	Number	Column Percent	Number	Column Percent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	28,160,000	100.00%	21,000,000	100.00%	5,557,000	100.00%	1,603,000	100.00%
U.S. Citizen	24,360,000	86.51%	17,900,000	85.24%	5,057,000	91.00%	1,398,000	87.21%
Non U.S. Citizen	3,527,000	12.52%	2,974,000	14.16%	357,000	6.42%	196,000	12.23%
Missing	279,000	0.99%	126,000	0.60%	144,000	2.59%	8,600	0.54%

Source: 2015 Nonemployer data and 2015 Census Numident.

Note: This table presents firm-owner level citizenship information by firm type. Owners of sole proprietorships are identified from Form 1040. Owners of partnerships and S-corps are identified from Schedule K-1. Citizenship information is obtained from the Census Numident. Only person owners are included in this table (i.e. firms that own other firms are excluded, as they cannot be matched to demographic information). Individuals appear once for each firm they own; individuals who own multiple types of firms appear in multiple categories. Missing indicates that an individual did not match to the Numident. For all individuals who are missing citizenship in the Numident, we assign the status "U.S. Citizen" (see text for more information).

Figures

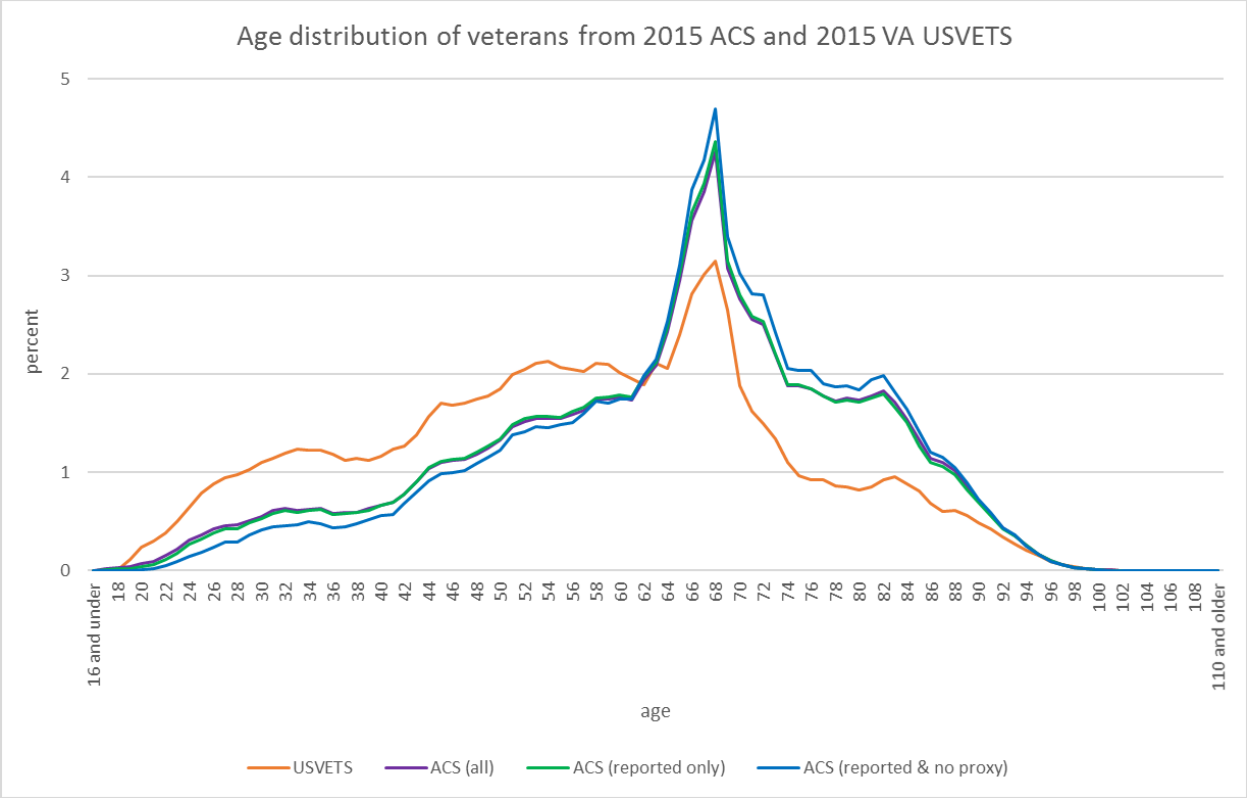
Figure 1: 2015 Nonemployer Universe



Source: 2015 Nonemployer Statistics publication.

Appendix Figures

Figure A. 1:



Source: 2015 ACS and 2015 VA USVETS.
Note: The USVETS series in the figure above excludes veterans known to the VA to be dead as well as veterans with missing date of birth (who are mostly dead according to the VA).

Appendix 1: Consolidated Surveys

The SBO was the only comprehensive source of information in the U.S. on employer and nonemployer businesses by the sex, race, ethnicity and veteran status of the business owner. It was a firm-level survey conducted every 5 years as part of the quinquennial Economic Census in years ending in a “2” or a “7”. Its sample frame included all nonfarm businesses with annual receipts of one thousand dollars or more. The ASE was conducted annually for survey years 2014 through 2016, and only covered employer businesses. Like the SBO, it provided information on selected economic and demographic characteristics for businesses and business owners by sex, ethnicity, race, and veteran status, but it also included additional questions on entrepreneurs' access to capital and financial barriers. Furthermore, it also added a rotating module meant to capture relevant issues facing businesses, such innovation and research and development (2014 ASE), management and business practices (2015 ASE), and business advice and planning (2016 ASE). These modules were designed to collect detailed information in periodic measurements within the same survey program with the same business sample of variables. The BRDIS-M was conducted in 2016 and was the primary source of information on research and development expenditures and R&D employees of for-profit, nonfarm businesses with fewer than 10 employees operating in the United States.

Appendix 2: Administrative Records Composite File

For the assignment of race and ethnicity, we are exploring the potential utility of using an administrative records and third party data (AR) race and Hispanic origin composite in addition to the previous Census and ACS responses in the PCR data file. The AR composite includes federal, state, and third party data files: PCR data (Census 2000 and 2010 and ACS 2001-2015), the Social Security Administration's Numerical Identification File, three different files from the Department of Housing and Urban Development, the Center for Medicare and Medicaid Services Medicare Enrollment Database, the Indian Health Service Patient Registration System, Temporary Assistance for Needy Families, Texas Supplemental Nutrition Assistance Program data, the Medicaid Statistical Information System, and four third party files.

Race and Hispanic origin responses in the AR composite are assigned using a set of business rules that differs from those used in the PCR file (as explained in the Methodology Section). In the AR composite, if there was no discrepancy in a person's Hispanic origin or race responses across administrative files, then the AR composite rules assign that race or Hispanic origin to the individual. If there were differences in a person's Hispanic origin responses across administrative files then the AR composite rules assign a 'Hispanic' response to the individual. If race responses were discrepant across files, a single race was assigned with preference given to smaller race groups according to their share of the total 2010 Census population distribution – Native Hawaiian or Other Pacific Islander (NHPI) alone, American Indian or Alaska Native (AIAN) alone, Two or More Races, Asian alone, Some Other Race (SOR) alone, Black alone, and White alone . (Ennis et al. 2015).

Appendix *Table A. 1* shows match rates for the nonemployer firm-owner pair universe linked to the AR composite and availability of race and Hispanic origin responses. There is a high match rate between nonemployer firm-owner pairs and the AR composite – overall as well as by LFO. About 97 percent of all firm-owner pairs can be linked to the AR composite data. By LFO, the match rates are about 97 percent for sole proprietorships, and approximately 97 and 99 percent for partnerships and S-corps respectively. This is higher than the match rate for the PCR data (between 93 and 96 percent for each type of LFO). We also see higher rates of availability of race and Hispanic origin responses using the AR composite (91.5 and 96.3 percent overall) relative to the PCR data (89.9 and 94.6 percent overall).

At the firm-owner level, Appendix Table A. 2 shows the distribution of race responses for all firm-owner pairs after the match to the AR composite data. The table also shows the percentage point difference between the two distributions. Firm-owner pairs with a match in the AR composite data and those without a match are included in the distribution. Compared to the match with PCR data, the match to the AR composite has 1.4 percentage points fewer missing race responses. Overall, the race distribution is similar between the two data sources with all race groups within a percentage point of the total share, and this distribution is similar across all legal forms of organization. While the racial distribution is similar and the percentage point difference small, the percent change increase for small race groups is high. More than three and a half times as many firm-owner pairs are assigned as AIAN using the AR composite (178,000) as are using the PCR data (50,000), and there are nearly 50 percent more assigned as NHPI using the AR composite (33,500) than are using the PCR data (24,500).

Appendix Table A.3 presents the distribution of race responses for firm-owner pairs matched to the 2015 PCR data and firm-owner pairs matched to the AR composite with the missing race responses removed from the distribution. The table again shows percentage point differences between the two distributions, which is similar and even smaller in magnitude than those seen in Appendix Table A. 2.

Appendix Table A.4 shows the Hispanic origin response distribution for the firm-owner pairs matched to the 2015 PCR data and firm-owner pairs matched to the AR composite, as well as the differences in the distributions. As with race, there are fewer missing Hispanic origin responses using the AR composite (3.7 percent) than there are using the 2015 PCR data (6.5 percent). Overall, the AR composite assigns more Hispanic responses (13.7 percent) to the nonemployer data than does the PCR data (10.4 percent). We see this across legal form of organization with about 3 to 4 percentage points more assigned to Hispanic for sole proprietorships, partnerships, and S-corps. Again, while the percentage point difference is relatively small, the percent change for Hispanics is more substantial. Using the AR composite, there are close to one-third more owners assigned a Hispanic response (3,844,000) than there are using the PCR data (2,920,000).

Appendix Table A.5 presents the distribution of Hispanic origin responses for firm-owner pairs matched to the 2015 PCR data and firm-owner pairs matched to the AR composite with the missing race responses removed from the distribution as well as the percentage point difference between the two distributions. Overall, the patterns observed in Appendix Table A.5 are similar to the patterns seen in Appendix Table A.4.

The availability of race and Hispanic origin responses in the AR composite for unique business owners according to legal form of organization is shown in Appendix Table A.6. As with

the PCR data, the AR composite has the highest match rate to owners of partnerships and S-corps (100.0), but whereas PCR data had the lowest match rate to sole proprietorships, the AR composite has the lowest match rate with owners of partnerships (95.8 percent). The AR composite has the highest availability of race and Hispanic origin for owners of partnerships and S-corps as well. Whereas the match rate is lowest for partnerships, race is least available for owners of sole proprietorships (89.7 percent). Hispanic origin is least available for owners of partnerships (95.0 percent).

The race distribution using the AR composite for unique business owners according to legal form of organization is shown in Appendix Table A.7. Again we see variation in the distribution of race responses according to legal form of organization when race is assigned using the AR composite. As with the distribution for all firm-owner pairs, and as when race is assigned using the PCR data, the AR composite assigns sole proprietorships the lowest share of White alone (70.2 percent) and the highest share of Black alone responses (11.8 percent). Whereas the PCR data assigned the highest share of White alone and the lowest share to owners of partnerships and S-corps, the AR data assigns the highest share of White alone (89.2 percent) to owners of sole proprietorships, partnerships, and S-corps. The AR composite assigns the lowest share of Black alone (1.2 percent) to owners of partnerships and S-corps, similar to the PCR data.

Appendix Table A.8 presents the Hispanic origin distribution for unique business owners by legal form of organization when the AR composite is used to assign Hispanic origin. As with the distribution for all firm-owner pairs, and like the PCR data, sole proprietorships have the highest share of unique business owners with a Hispanic response (16.4 percent) and the lowest share of non-Hispanic responses (79.8 percent). And while unique owners of partnerships and S-corps have

the lowest share of Hispanic responses when using the PCR data, the AR composite assigns the lowest share of Hispanic responses (5.8 percent) to owners of sole proprietorships, partnerships, and S-corps. The AR composite assigns the highest share of non-Hispanic responses (93.1 percent) to owners of partnerships and S-corps.

Turning to firm-level results, Appendix Table A.9 shows results for assignment of race responses at the firm-level by the AR composite. The distributions across assignment rules and across legal forms of organization are similar to those seen in Table 26for the PCR data. There is a slightly higher percentage of White-owned and Black-owned firms using AR composite responses compared to PCR responses but this may result from the lower percentage of firms without any owners linkable to the AR composite. Also note that the share of AIAN-owned firms is higher using the AR composite. This is a result of the inclusion in the AR composite of the Indian Health Services data.

Appendix Table A.10 shows firm-level Hispanic origin assignment using the AR composite. Compared to the distribution using the PCR data, the AR composite assigns more firms as being Hispanic-owned (15.0 percent) relative to the PCR data (11.5 percent). The share of non-Hispanic owners using the AR composite is more similar and within a percentage point of the share using the PCR data. Both of these trends hold true across method of assignment and across legal form of organization. More firms can be linked to the AR composite overall with only 2.6 percent missing race because no firm owners could be linked to the composite, compared to 6.5 percent of firms that could not be linked to the PCR data. These additional linkages provided by the AR data may contribute to the higher share of Hispanic-owned firms.

Appendix Table A.11 compares the aggregate race distribution of firms as assigned by the PCR data and the AR composite with the 2002, 2007, and 2012 SBO. Comparing the 2015 distribution with previous years, we see that the shares of White-owned, Asian-owned, and NHPI-owned firms is within the range for prior years of the SBO. The AR composite assigns more Black-owned firms and more AIAN-owned firms compared to the SBO.

Similar comparisons are presented in Appendix Table A.12 for Hispanic origin. Compared to both the SBO and the PCR data, the AR composite assigns more firms as Hispanic-owned (15.6 percent) and fewer firms as non-Hispanic-owned (84 percent). The AR composite, like the PCR data, assigns 0.2 percent of firms as being equally owned.

As we continue to research the use of previous Census records and administrative records for the assignment of race and ethnicity to nonemployer business data, we will evaluate the appropriateness and implications of using an AR composite. We will analyze the AR data to measure any biases that may be introduced by the administrative records as well as the long-term feasibility of relying on an AR composite. The composite used in our research includes Third Party Data that are not obtained and updated routinely, and so the utility of those files will decrease over time. Other data are acquired using data sharing agreements with other federal agencies, and there is no guarantee that such agreements may change over time. For example, if the Indian Health Services data were not available, this would impact the assignment of AIAN using the AR composite. As we continue with our research, we will strive to determine the most appropriate data to assign race and Hispanic origin to nonemployer business data that achieve the high level of quality associated with past nonemployer Census products.

Appendix 3: DOD's Defense Enrollment Eligibility Reporting System (DEERS) Database

As described in the Data Section above, individuals who are currently serving on active military duty are not considered veterans per the VA's definition. But these individuals, if self-identified as such, are considered veterans by the SBO and the ABS surveys. Similarly, some individuals in the National Guard/Reserve component are not included in the USVETS database limiting our ability to accurately identify these military personnel.

We believe that both of these groups of individuals can be identified from the DOD's Defense Enrollment Eligibility Reporting System (DEERS) database. The DOD DEERS database includes data on individuals who are currently serving on active military duty, and is also likely to allow us to more accurately and completely identify individuals in the National Guard/Reserve component relative to the VA USVETS database. Our research indicates that the VA USVETS and the DOD DEERS databases would complement each other allowing for the most comprehensive, and consistent with the surveys, identification of veterans among nonemployer firm owners. Each of these two administrative records sources of veteran status covers different types of veterans. The VA USVETS database covers individuals who served on active military duty in the past. The DOD DEERS database covers individuals serving on active military duty now and those serving in the National Guard/Reserve component (excluding the Coast Guard Reserve component).

While the USVETS data contain information on individuals in the National Guard/Reserve component, not everyone serving in the reserve components of the U.S. military is captured by this administrative data source. In theory, the USVETS database should include only ever-activated, meaning they served a qualifying active duty period in the past, veterans who are serving in the

National Guard/Reserve component but based on our communication with the VA we learned that this group of veterans is not accurately captured in the current version of the USVETS database.[87] The USVETS database pulls in data on veterans in the National Guard/Reserve component from the DOD DEERS database. The DOD DEERS database is unable to distinguish between individuals who are ever-activated versus those who are never-activated and serving in the National Guard/Reserve component. Until the DOD resolves this data issue, the USVETS database will continue to misidentify some individuals serving in the National Guard/Reserve component. For the purposes of this project, however, the distinction between the ever-activated and the never-activated individuals in the National Guard/Reserve component is actually not crucial because the SBO and the ABS surveys consider anyone serving in the National Guard/Reserve component, regardless of their past activation status, to be veterans. Our research suggests that in order to accurately identify individuals serving in the National Guard/Reserve component, both activated and not, and to be consistent with the surveys' concept of a veteran, the most likely best source of this information is the DOD DEERS database.

Both the DEERS and the USVETS databases would allow us to identify individuals serving in the Coast Guard Reserve component who have been activated at least once in the past (and hence are considered veterans by the VA). But neither the DEERS nor the USVETS data cover never-activated individuals in the Coast Guard Reserve component. As these veterans represent a fraction of one percent of all military servicemen, this is a relatively minor data limitation. We will continue

[87] It is also unclear whether these veterans are under- or over-counted in the current version of the USVETS database.

to explore the DOD DEERS database as an additional and complementary source of veteran information.

As this project progresses and more detailed characteristics related to nonemployer firm owners' military service are considered, we briefly explored the potential to rely on either the USVETS or the DEERS data for assignment of service-connected disability information to these owners. We find that a concept of a service-connected disability is more broadly defined in the SBO and the ABS surveys than in either the USVETS or the DEERS databases.[88] The USVETS data identifies veterans with a service-connected disability only if they are being compensated by the VA for their military service disability. On the other hand, the DEERS data identifies veterans with a service-connected disability only if they are being compensated by the DOD for their disability. But the DOD and the VA have different definitions of a service-connected disability. We will continue to explore conceptual differences across potential data sources of this disability information.

[88] The survey asks "Is Owner X disabled as the result of illness or injury incurred or aggravated during military service?"

Appendix 4: SBO & ABS military service questions

MILITARY SERVICE

Has Owner X ever served in any branch of the U.S. Armed Forces, including the Coast Guard, the National Guard, or Reserve component of any service branch?

☐ Yes

☐ No – SKIP to INITIAL ACQUISITION

MILITARY SERVICE DISABILITY

Is Owner X disabled as the result of illness or injury incurred or aggravated during military service?

☐ Yes

☐ No

OTHER MILITARY SERVICE

Do any of the following characteristics describe Owner X's military service? *Select all that apply.*

☐ Served on active duty military service, not including training for the Reserves or National Guard

☐ Served on active duty military service after September 11, 2001

☐ Served on active duty military service in 2017

☐ Served in the National Guard or as a reservist of any branch of the U.S. Armed Forces in 2017

☐ None of the above

Appendix 5: Quality Assessment of VA's USVETS

Because the USVETS database is the primary data source we are considering to assign veteran status information to nonemployers, it is important to conduct some quality checks relative to other data sources on veterans. In this section, we briefly discuss our findings regarding the quality of these data. In general, the USVETS data are less accurate for older and healthier veterans. Because of the 1973 fire at the National Personnel Records Center that destroyed approximately 16 to 18 million official military personnel files older veterans are likely to be undercounted in the USVETS database.[89] As older veterans die, this limitation of the USVETS data will gradually decrease over time. Due to the USVETS reliance on various administrative records sources, veterans who have never used the VA medical facilities or who have never enrolled in the VA benefit programs are not captured by the USVETS data. As a result, healthier veterans may be undercounted in the current version of the USVETS database. As the USVETS database is updated with data from additional administrative and commercial sources, this limitation should also decrease over time.

We evaluate the quality of the USVETS data by linking it to the 2015 American Community Survey (ACS) and the 2012 SBO data at the individual level. When we link veteran status responses of first owners of sole proprietorships from the 2012 SBO to their veteran status information in the USVETS data, we find that approximately 44 percent of the individuals identified as veterans by the SBO are found in the USVETS database (see Appendix Table A.13). [90]

[89] See more information about this fire and its impact here: https://www.archives.gov/personnel-records-center/fire-1973

[90] The sample consists of reported cases of in-scope firms for owner 1 of nonemployer sole proprietors. As SBO includes PIKs for owner 1 only, we are not able to examine other owners' veteran status responses.

While low, this match rate is not surprising given the different time periods covered by the survey and the administrative records dataset, and the potential for proxy and imputed veteran status responses in the survey. Of these owners in the 2012 SBO who reported being non-veterans, their responses matched to the USVETS data over 95 percent of the time.

When we link individual veteran status responses from the 2015 ACS to the USVETS data, we find a much higher match rate. Specifically, we find that approximately 76 percent of individuals identified as veterans in the ACS based on the VA definition of a veteran[91] are found in the USVETS database (see Appendix Table A.4). This match rate increases slightly to approximately 79 percent when we exclude imputed and proxy responses from the ACS sample (see Appendix Tables A.15 and A.16). This discrepancy between the ACS and the USVETS data is likely due to two factors. First, the VA cannot accurately distinguish between ever- and never-activated reservists in the current version of the USVETS database. And, second, the ACS respondents who served in the past and continue to serve in the present are required to select only one response - either "now on active duty" or "on active duty in the past, but not now." If most such respondents select "now on active duty" answer choice, then the response rate to the "active duty in the past, but not now" category, which is conceptually consistent with the USVETS database, is understated.

We also examine the age distribution of the ACS veteran respondents relative to the age distribution of veterans in the USVETS data. As expected, younger veterans are better represented in the USVETS than older veterans (see Appendix Figure 1).

[91] This corresponds to the following ACS response to the military service question "on active duty in the past, but not now."