Preliminary and Incomplete Draft
Do Not Cite Without Permission

Reconciling Household and Administrative Measures of Self-Employment and Entrepreneurship*

August 2006

Melissa Bjelland
Cornell University
U.S. Bureau of Labor Statistics
U.S. Census Bureau

John Haltiwanger
University of Maryland
National Bureau of Economic Research
U.S. Census Bureau

Kristin Sandusky
U.S. Census Bureau

James Spletzer
U.S. Bureau of Labor Statistics

* This document reports the results of research and analysis undertaken in part by U.S. Census Bureau staff. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854, and the Alfred P. Sloan Foundation. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The paper has been screened to ensure that it does not disclose any confidential information. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov http://lehd.dsd.census.gov).

I. Introduction

Have changes in the economy blurred the boundaries of the population of self-employed, a large and historically difficult to quantify segment of the workforce? To date, household-based surveys such as the Current Population Survey have provided the leading source of information on the self-employed. A non-trivial fraction of the U.S. labor force has ties to self-employment: in a calendar year, roughly seven percent of the working population is self-employed at their main job, with close to nine percent receiving some income derived from self-employment when secondary sources of earnings are included. These self-employed individuals operate nearly thirty percent of firms with employees and eighty percent of firms without employees. While numbers alone make this group of interest, topics such as the noted growth of small businesses and employment discrepancies between household and business data have drawn attention recently to the population of self-employed workers. It is unknown how well respondent reports of self-employment align with information from administrative sources and how disagreements may have changed over time. The measurement difficulties for this group of workers and businesses have been compounded by a recent (perhaps technology-induced) increase in the prevalence of small, self-directed sources of additional household income and outsourcing (contracting out). These practices have made the standard survey question "Are you self-employed?" less straightforward to answer.

Self-employment can be measured independently from surveys of households, surveys of businesses, and administrative records. While a broad concept of self-employment is present in each, potentially important subtle differences exist in defining and interpreting self-employment across these household and business data sources. Many individuals who are self-employed have multiple sources of income that include wage and salary earnings, so household-based questionnaires that focus exclusively upon a worker's main job may omit evidence of self-employment activity for those who hold secondary positions. As a result, person-level statistics on employment classification might be subject to inaccuracies. These problems cannot be simply rectified by utilizing administrative data that capture all job matches, since such records bring along their own issues relating to scope and measurement. Without a single data source known to be unequivocally comprehensive and reliable, it is apparent that characterizing the group of self-employed individuals involves untangling misclassifications and errors in self-employment status and earnings. To obtain a better understanding of the level and dynamics of this population of workers, we analyze self-employment at the aggregate and micro level by merging together survey and administrative data.

In particular, this paper integrates and compares three sources of self-employment data: the Current Population Survey (CPS), the Social Security Administration's Detailed Earnings Record (DER), and the Census Bureau's Business Register (BR). The Current Population Survey is a household-based survey of the U.S. population with respondent-reported information on employment and earnings. The Detailed Earnings Record contains complete earnings history information originating in 1978 for CPS

respondents obtained from Social Security administrative data. Like the CPS, the DER covers both wage and salary workers as well as the self-employed. The Business Register provides universal coverage of all employer and non-employer sole proprietorships, partnerships, and corporations from various administrative and survey data sources. The Business Register is important in this context because it provides commonly-used statistics on U.S. businesses including the contribution of sole proprietors. It is, in a related way, the sample frame for surveys of business owners such as the Survey of Business Owners (SBO) and Surveys of Minority Owned and Women Owned Businesses (SMOBE/SWOBE). For this study, we restrict the set of Business Register records to those that can be merged, via a person identifier, into the CPS and DER. From these linked records, we construct overall self-employment rates and self-employment earnings measures from each source. This union of files permits the comparison of the status of self-employment across data sources. It also allows the difference in self-employment earnings across files to be examined. Our microanalysis- with a base focus on CPS respondents integrated with administrative data- permits us to characterize the propensity for mismatch as a function of worker and job characteristics.

This paper is organized as follows. Section II reviews the current literature on self-employment measurement, highlights the role that self-employed workers may play in improving our understanding of small business dynamics, and emphasizes how a clear account of their numbers may help reconcile the employment disagreement between household and business data. Section III describes the integrated dataset we use to form our comparison. In particular, it addresses the manner in which each of the three data files defines self-employment and identifies how these sources may or may not overlap. In this section, we also report self-employment rates and the distribution of self-employment earnings from each source. In section IV, we explore micro-level differences in self-employment classification between these data. Focusing on the set of workers identified as self-employed in at least two of the three sources, we characterize the distribution of differences in self-employment earnings. We also more closely examine the misclassification in self-employment and the disparity in self-employment earnings by worker and firm traits. Section V concludes.

II. Background

In this section, we clarify our definition of self-employment, we provide an overview of the various sources of information on the self-employed produced by the US statistical system, and we summarize the strengths and weaknesses of these sources for accounting purposes as well as for alternative types of statistical analysis. We also provide details on the findings of previous measurement-related research. Lastly, we present examples of conclusions drawn from comparisons of the US workforce as portrayed by differing sources from studies of the populations of the self-employed and highlight how these conclusions may be influenced by misclassification of the self-employed.

Conceptually, what do we mean by self-employment and what groups and activities should statistics on self-employment cover? One perspective on self-employment focuses on the person and their employment status, whereas an alternative perspective

highlights the self-employed entity as a business or an entrepreneur. From either perspective, one critical feature of data on self-employment is that it should permit the enumeration of self-employed workers at a given point in time. A person is considered to be self-employed if the individual generates revenue from some market-based activity and earns this revenue as an independent entity and not as an employee of some existing business.

Given the definition above, a number of measurement and conceptual issues immediately arise. For example, do workers who are independent contractors to a specific business for an extended period of time fall into the category of self-employed? For tax reporting purposes, these individuals may be counted as independent business entities, yet behaviorally, they may perform the same activities and have the same relationship with the contracting business as do traditional wage and salary employees. While the decision to include (or exclude) these workers in published statistics may be subjective, the question of whether dissimilar data sources have the capacity to accommodate either alternative currently remains unanswered. Are these workers included in records from some data files while they lack a universal presence in statistics published from other sources? More importantly, are the existing data on the self-employed detailed enough to permit an evaluation of these potential imbalances or even to empirically distinguish these workers from the more traditional self-employed?

The definition of self-employment we have presented is sufficiently ambiguous that it also is unclear whether individuals who engage in self-employment as a hobby as opposed to a primary source of income should be counted. It is of interest to be able to make this distinction since hobby activities may transition into primary sources of income when the hobby grows sufficiently large or when the "main job" is terminated due to an event such as layoff or retirement. To date, little is known about the relative abilities of alternative data sources to capture this type of hobby-related business activity.

The importance of young and small businesses in job creation, innovation, and productivity growth has received much attention. Although the vast majority of these types of business owners are sole proprietors (with and without employees) who we would expect to find classified as self-employed in any source of information, data sources that either omit or poorly measure the early stages of these smaller ventures limit our ability to fully characterize the origins of significant self-directed economic activity. Recent evidence suggests that these self-employed may have even larger implications upon U.S. business dynamics than was previously thought. Figure 1 (drawn from Davis et. al. (2005)) indicates that a significant number of new employer businesses have a pre-history as non-employer businesses, most of which are sole proprietors and partnerships. These findings suggest that the accurate measurement of U.S. employer business startups also depends upon the precise tracking of the self-employed over time. An evaluation of how well alternative sources measure this activity has not yet been performed.

Household-based surveys, such as the Current Population Survey (including the Basic Monthly and March Supplement files) and the Survey of Income and Program Participation have provided the primary source of information on rates of self-

employment and demographic detail of the self-employed. Advantages of using these household-based surveys to characterize the self-employed population include nationally-representative samples, much detail of the business owner, the availability of lengthy time series, and the presence of wage and salary workers as a comparison group. In addition, these surveys facilitate the discernment of self-employment that is linked to the primary source of income from activity that is supplemental. Panel data on individuals, such as the National Longitudinal Survey of Youth, the Panel Survey of Income Dynamics, and the SIPP provide longitudinal information on self-employment and permit the measurement of transitions into and out of self-employment. In addition, many of these longitudinal datasets provide detailed information about a respondent's business, such as ownership, work, and earnings histories. In this study, we focus on one example from the class of household data sources – the CPS.

Clear knowledge of self-employment is key for the reliable measurement of employment status, employment fluctuations, the number of businesses, and business dynamics. On the employment side, the role of self-employment has often been raised as being one of the factors underlying the differences between household and establishment employment statistics. In principle, precise accounting of self-employment status should correct the discrepancies that occur when self-employed workers are enumerated in the household surveys but not in the establishment statistics (which are based on a mixture of survey and administrative data). However, even after carefully controlling for the number of *measured* self-employed, inconsistencies continue to persist between household- and establishment-based statistics. Figure 2 illustrates the patterns of CES non-farm payroll employment (from establishment data) and CPS non-farm employment (from household data) after the CPS has been adjusted to make it look more similar to the CES (see Bowler and Morisi (2006) for a description of this adjustment – removing the self-employed is the largest of the many adjustments). This graph illustrates that a substantial gap remains to be explained in the time series of employment rates. One possibility for this imprecision is the misclassification of self-employment status.[1]

As noted, the household and administrative sources of self-employment data each have their own advantages. Each provides a unique perspective (to varying degrees) of self-employment and entrepreneurial activity. Yet much overlap exists across sources as well in terms of covered activity. It is an established fact that even holding constant the subpopulation of interest, self-employment activity varies across household and business data sources. As an example, Boden and Nucci (1995) explore self-employment counts for different groups of workers using alternative samples that have been restricted to reflect the same base population of self-employed. They compare measures of self-employment from the CPS March Supplement file to identical measures found within the

[1] The Federal Economics Statistics Advisory Committee (FESAC) has recently issued a report and recommendation that the Bureau of Labor Statistics explore the potential role of self-employment misclassification in accounting for the trend and cyclical fluctuations in the difference between establishment- and household-based statistics on employment. Similarly, House Resolution 14 in the 109th Congress (the Dreier Boehner resolution -- see http://www.dreier.house.gov/releases/pr010405c.htm) requests that the BLS explore the role of self-employment (and other factors) for this statistical discrepancy.

Characteristics of Business Owners data. Even after restricting each survey to industries common to both, and after excluding the incorporated self-employed, substantial differences in the level of self-employed workers remain. These gaps become even more pronounced when the comparison is made within cells of workers with common traits. Only self-employment counts across geographic and educational groups are found to be similar.

Similarly, Fairlie (2005) finds a much larger number of businesses from sole proprietors, partnerships, and sub-chapter S corporations from counts of businesses from the Survey of Business Owners and SMOBE/SWOBE Census data than emerge from the Current Population Survey. Fairlie additionally notes the popularized argument that entrepreneurship provides a route out of poverty and an alternative to unemployment or discrimination in the labor market and notes as well that this argument has led to policies and programs designed to promote entrepreneurship and business ownership among disadvantaged groups. In evaluating factors related to discrepancies in transitions into and out of self-employment across these demographic groups, relatively little work has been done to evaluate the possible role that data quality issues may play in contributing to these across-group differences. Although Fairlie finds that disadvantaged groups have substantially lower rates of transition into self-employment and higher rates of transition out of self-employment than do other groups, these discrepancies cannot be completely accounted for by across-group variation in observable traits, such as asset levels.

In a paper closely related to the analysis in the current paper, Roemer (2002) links CPS records to the DER. His primary focus is to account for differences in the patterns of earnings for wage and salary workers across survey and administrative data.[2] However, Roemer discovers that "a non-trivial difference in the definition of 'self-employment' exists between the March monthly CPS and administrative systems." Specifically, he finds that workers classified as wage and salary workers in the CPS but as self-employed in the DER are found disproportionately in occupations dominated by independent contractors such as truck drivers and construction workers. These are both occupations that have a higher share of workers from the same disadvantaged groups that Fairlie examined. Thus, it is possible that jobs less easily classified by respondents as self-employment are concentrated in occupations held disproportionately by disadvantaged groups. Roemer's findings suggest that issues related to the quality of survey data may be important for understanding the variation in the data across such groups.

Our value-added relative to the above literature is that we integrate the CPS, the DER, and the BR to comprehensively explore the nature of the discrepancies in defining the population of self-employed workers at both the aggregate and micro level. We exploit our micro-level linkages to compare the ability of each data source to provide coverage of groups known to be difficult to measure, such as contractors and smaller entrepreneurial ventures. We now turn to a detailed description of these data sources.

[2] A related paper that explores earnings differences across survey and administrative data for the SIPP is Abowd and Stinson (2003).

III. Three Sources of Self-Employment Data

We create measures of self-employment from three different datasets: the Annual Demographic Survey to the Current Population Survey (often called the March Supplement); the Social Security Administration's Detailed Earnings Records; and the Census Bureau's Business Register. The March Supplement is a well-known household survey, whereas the latter two data sources are constructed from administrative data and may be less familiar to economists. In this section, we describe these three data sources and the measures of self-employment available within them, and we present some simple macro-level comparisons. In the next section of this paper, we present results from merging the three files at the microdata level.

IIIA. CPS Annual Social and Economic Supplement

The Current Population Survey (CPS) is a monthly survey of approximately 60,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The data from this survey provide a comprehensive body of information on the employment and unemployment experience of the Nation's population, classified by age, gender, race, and a variety of other characteristics. The survey has been conducted for more than 50 years, and the CPS microdata that are made available to researchers have helped create the empirical foundations for much of modern labor economics. The CPS collects information on the labor force status of the civilian non-institutional population 15 years of age and older, although labor force estimates are reported only for those 16 and older. Respondents are assured that all information obtained is completely confidential and is used only for the purpose of statistical analyses.

The Annual Demographic Survey (ADS) to the CPS is the primary source of detailed information on income and labor force participation in the United States. The ADS collects information on the employment status of persons 15 years old and over during the previous calendar year. Those who are in an active rotation of the CPS in March are respondents to the ADS questionnaire. Additionally, the members of roughly 3,000 households of Hispanic origin who were interviewed the prior November are also contacted in March and their answers are included as part of the supplement. Beginning in 2001, the March ADS was further expanded to one-quarter of the respondents to the February and April surveys who were out of the CPS rotation in the month of March.[3] Each year, the Bureau of Labor Statistics and the Bureau of the Census issue numerous publications based on the March Supplement. For example, the survey is used to generate the annual Population Profile of the United States, reports on geographical mobility and educational attainment, and detailed analyses of money income and poverty status.

[3] Households selected for the CPS sample are entered into a 16-month rotation pattern in which they are contacted in each of the first 4 months, temporarily leave the survey for 8 months, and then are recontacted in each of the final 4 months (often called a 4-8-4 rotation pattern). For more details on the March ADS sample, see http://www.bls.census.gov/cps/ads/data_dissem_letterng.htm.

We use the March Supplement microdata to define whether an individual was self-employed during the previous year. According to the CPS webpage, self-employed persons are those who work for profit or fees in their own business, profession or trade, or operate a farm.[4] Many academic studies have analyzed self-employment from the Annual Demographic Survey – see, for example, Devine (1994), Fairlie and Meyer (1996), and Carrington, McCue, and Pierce (2000).

We define an individual in the ADS as self-employed in two ways. First, and most common, an individual is self-employed in the previous calendar year if he describes the longest job held during the previous year as "self-employed or working in a family business." We use this definition from the CPS to identify self-employment as the "main job." Individuals must also state that the business was not incorporated and must be identified as the owner of the business. The second way to define an individual as self-employed in the ADS data is if their longest job held during the previous year was not "self-employed or working in a family business" but they report some self-employment income from other work during the previous calendar year. The exact wording of the supplement questions and the responses necessary to define an individual as self-employed are given in Exhibit 1 of the appendix.[5] We derive a broader "all jobs" measure of self-employment by including both those for whom self-employment is the main job and those who have some self-employment income in the prior year.

One aspect of our Annual Demographic Survey measure of self-employment warrants mention. Individuals in the survey who report that they were "self-employed or working in a family business" are then asked whether the business was incorporated. We have chosen to exclude these incorporated self-employed workers from our definition of self-employment for several reasons. Primarily, the official self-employment statistics from BLS do not include this group (see Hipple (2004)). The logic behind this exclusion is that the individuals who own corporations can often also be employees of their own corporation and then be classified as wage and salary workers. Secondly, our measures of self-employment from the administrative data exclude the incorporated business owners, and we want, as best as possible, similar conceptual definitions for our comparison of self-employment across datasets.[6]

---

[4] See http://www.bls.census.gov/cps/ads/1995/sglosary.htm.

[5] In a data appendix that is available upon request, we describe how we use the actual variables and data from the ADS to determine self-employed individuals.

[6] In considering the types of business owners, it is important to emphasize that sole proprietors, partnerships and corporations are all distinct and mutually exclusive legal forms of organization. For the SSA DER, we have information on sole proprietors and partnerships. We have this since we have self-employment income as reported on Schedule SE of individual tax returns for all matched CPS-DER cases. Schedule SE includes self-employment income from Schedule C (sole proprietors), partnerships and farm businesses (we exclude farm businesses from our analysis). Note that business owners that are corporations are not subject to self-employment tax and do not file Schedule SE. Business owners that are corporations file corporate income taxes and distributions (e.g. dividends) from the corporation to the business owner are subject to individual income tax but not self-employment tax. For the BR, we have sole proprietors only for this analysis. The reason that we are restricted to sole proprietors for this analysis is that we have personal identifier keys (PIKs) for sole proprietors and thus can integrate the sole proprietor data into the CPS-DER matched data. We note that the BR includes partnerships and corporations but at this point we do not have PIKs for the business owners for those types of businesses. In future work, we

Vitally important to our analysis in this paper is the ability to match individuals in the ADS to various administrative datasets at the U.S. Census Bureau. Approximately 70-80% (the exact number varies by year) of individuals in the supplement have what is called a Personal Identification Key (PIK). The PIK is a unique internal Census Bureau identification number for the individual.[7] We use the PIK to link the CPS records with the administrative records in the Social Security Administration's Detailed Earnings Records and in the Business Register– these linkages enable the comparisons reported later in this paper. Even though the percentage of records with a PIK is high, it is far from 100 percent and there are naturally concerns about sample selection. In our analysis below, we use propensity score methods for adjusting sample weights to deal with this issue. The methods are described below.

We further restrict our CPS sample to workers in the non-agricultural sector, aged 25-65, and who are not armed forces personnel in their longest job held in the year prior to the survey. The reason for removing workers in the farming and fishing industries and from occupations in the armed forces is to make our comparisons to the administrative data as conceptually clean as possible. As will become evident when we describe the administrative data, farmers submit different tax forms than do self-employed persons in the non-agricultural sector, thus making their administrative data inherently different. Furthermore, to be consistent with the bulk of the academic literature, we believe that this initial comparison of self-employment across the three datasets would benefit from focusing on the non-agricultural self-employed.

Descriptive statistics of self-employment in 1997 from the ADS are given in Table 1. These 1997 data are from the 1998 ADS, since the questionnaire asks about labor market activities in the previous year. The first column presents statistics from the full sample of the March Supplement. The second column presents statistics for the sample with a PIK. In this second column and for the remaining analysis throughout the paper, all numbers are generated using an adjusted version of CPS's March Supplement weight. This adjustment accounts for the records without PIKs that are omitted from our analysis. To perform this adjustment, each individual weight is multiplied by the inverse of the predicted probability that a PIK is available for the respondent. These probabilities are estimated using a logit model of the probability a PIK is provided as a function of age education, gender, race and work status. The use of direct propensity scores has been advocated by Kalton and Flores-Cervantes (2003) and used in programs like MEPS (see Wun et. al. (2004)).

---

plan to also explore the integration of self-employed partnerships and corporations from the BR into the analysis.

[7] For the ADS of the CPS, the Census Bureau collects SSNs of sampled people in selected years, and verifies these SSNs with the Social Security Administration. If a respondent doesn't know an SSN, or if it fails verification, the Census Bureau searches the SSA's database for it by name, date of birth and geography. If a respondent refuses to provide an SSN, the Census Bureau makes no attempt to find it. Validated SSNs link a sampled person to the DER. The resulting internally linked file uses as its person identifier a Personal Identification Key (PIK). The latter is an internally generated unique identifier and is the person identifier that is used to link administrative and survey data in this analysis. Note that an administrative records division at Census conducts the SSN validation and also assigns the PIK to a person.

In Table 1 and what follows, we define workers as employed in the CPS if they have positive earnings from any source. The restriction on positive earnings is limiting in that some self-employed workers have negative income (essentially negative profits from self-employment income) so that we are undercounting CPS self-employed workers by imposing this restriction. We make this restriction because one of our key administrative sources relies upon identification of self-employment income for workers with positive earnings above a threshold. Future drafts of the paper will investigate the role of self-employed workers with negative earnings.

In column 1 of Table 1, we see that there are about 7.5 million persons who are self-employed at their main job and about 9.6 million persons with some self-employment income in 1997. The self-employment rates in the sample with PIKs are quite similar to those derived from the full sample March Supplement. The self-employment rate for workers where self-employment is the main job is computed as the number of self-employed divided by the number of employed, and is about 6.8 percent for the full sample and about 6.6 for the sample with PIKs. For workers with any self-employment income, the corresponding self-employment rates are both about 8.7 percent.

Our comparison of self-employment across datasets will also focus on the earnings of the self-employed. Earnings of the self-employed are straightforward to measure in the March Supplement. As seen in Exhibit 1 of the appendix, individuals whose longest job during the previous year was self-employment are asked, "How much did you earn from this [self-employment job] before taxes and other deductions during the previous year?" For main job earnings statistics we use this measure as the measure of earnings. Individuals whose longest job during the previous year was not self-employment are asked whether they earned money from any other work during the previous year, and if yes, "How much did you earn from your own business after expenses?" Our measure of self-employment earnings during the previous year for workers with any self-employment income is the sum of the response to the latter question and, as appropriate, the question about the earnings on the longest job.

The mean and various distributional statistics for self-employment earnings are given in Table 1. The statistics for both the full sample and the sample with PIKs are quite similar. For main jobs, the mean for both the full sample and the sample with PIKs is just above $30,000 while for all jobs the mean for both samples is around 26,000. Not surprisingly, each of the percentiles for main jobs is substantially higher than for all self-employed jobs.[8] The tenth percentile for all jobs is quite low (around $300) which points to some of the challenges involved in measuring self-employment jobs and income. As will become clear below, this tenth percentile value falls below minimal thresholds for inclusion in some of the administrative data.

IIIB. Social Security Administration's Detailed Earnings Records (DER)

[8] For disclosure reasons we don't compute or report actual percentiles but rather all reported percentiles in this paper are based on the average of the five centiles centered on the percentile in question, rounded to the nearest $100.

The Social Security Administration maintains a detailed earnings history for all individuals. The following description is from http://www.ssa.gov/mystatement. An individual's employer is responsible for keeping a record of the wages paid throughout the year, withholding Social Security taxes, and submitting those taxes, plus a matching share, to the Internal Revenue Service. After the end of the year, an individual's employer sends the Social Security Administration a copy of form W-2, which shows the worker's name, Social Security number, and earnings. The Social Security Administration then records this information from the W-2 to the individual's detailed earnings history. In addition, SSA receives information from the IRS about self-employment earnings and related self-employment Social Security taxes. The Social Security Administration is required by law to send a Social Security Statement to eligible people – the updated Statement is sent each year about three months before a person's birthday. The Social Security Statement is a concise, easy-to-read personal record of the earnings on which a person has paid Social Security taxes during their working years and a summary of the estimated benefits that person and their family may receive as a result of those earnings.

The U.S. Census Bureau matches the March CPS microdata, by PIK, to the Social Security Administration's Detailed Earnings Record (DER). As stated above, the PIK is an internal Census Bureau individual-specific identification number, and all employment and earnings information associated with a PIK is completely confidential and is used only for the purpose of statistical analysis. We do not have access to the complete DER dataset; the DER records we have are contained with an exact-match file for the individuals in the March CPS (both employed and not employed).

Employment in the DER is defined by the presence of positive earnings as reported by the employer on the worker's W-2 form, or the presence of self-employment income. Self-employment in the DER is defined from line 4 of Schedule SE (self-employment tax) filed with the individual's IRS 1040 form. A simplified version of Schedule SE is given in Exhibit 2 of the appendix. Self-employment income derives from three sources: farms (individuals filing IRS schedule F), sole-proprietorships (individuals filing IRS schedule C), and partnerships (individuals filing IRS schedule K-1). The importance of having both sole-proprietorships and partnerships in the DER will become evident when we describe the self-employment measures available from the Census Bureau's Business Register.

As stated in the instructions to Schedule SE (and restated at the top of Exhibit 2 in the appendix), "you must pay Self-Employment tax if you had net earnings of $400 or more as a self-employed person." Technically, as seen on line 4 of Schedule SE, individuals only file this schedule if self-employment earnings are greater than $433.13 ($400 divided by 92.35%). In the DER microdata, some records with earnings below this threshold exist. We suspect that this is because some individuals with non-zero self-employment earnings file Schedule SE with their return even though they are not subject to the self-employment tax (and are not required to file this form).

Basic descriptive statistics about employment, self-employment, and self-employment earnings from the 1997 DER are given in Table 2. The sample here is those individuals with a PIK who are present in both the March CPS and the DER. We have applied the CPS March Supplement weights to DER persons in the matched March CPS - DER sample. In Table 2, we see that the DER identifies 10.6 million self-employed persons with any self-employment income and 7.6 million persons where self-employment is the main job in 1997. For the DER, the main job is defined as the job with the highest share of earnings. The self-employment rate, computed as the number of self-employed divided by the number of employed, is 7.2 percent for main jobs and 10.3 percent for all jobs. In comparing these statistics to Table 1, it is important to note that the definition of overall employment differs between the CPS and DER (this is a topic we discuss again below). Not only do the numerators of the self-employment rates differ in Tables 1 and 2, the denominators are also dissimilar (although both Tables 1 and 2 are based upon persons who are found in both the CPS and DER). With these caveats in mind, the DER yields somewhat higher self-employment rates than the CPS for both main jobs and for all jobs.

Self-employment earnings in the DER are measured from line 4 of the individual's Schedule SE, with the minimum level of self-employment earnings in principle at $433. The mean and various distributional statistics for self-employment earnings are given in Table 2. The mean for main jobs and all jobs are about $19,000 and $16,000, respectively. Comparing Tables 1 and 2, it is interesting that average earnings for self-employed are higher for both main jobs and all jobs in the CPS. Given the income threshold in the DER, the tenth percentile for all jobs is higher in the DER. All of the other percentiles for main jobs and all jobs are higher in the CPS than in the DER. It is clear from this that the distribution of self-employment income in the CPS lies to the right of the distribution of self-employment income in the DER.

IIIC. Census Bureau's Business Register

The U.S. Census Bureau's annual Business Register (BR) is the list of establishments the Bureau uses to develop the initial mailing list for the economic censuses and surveys. The BR provides universal coverage of businesses, with and without employees, that file payroll and income taxes with the IRS, and it contains data from several different sources: forms sent by each business to the IRS for tax purposes, or data taken from the economic censuses in census years or on the Company Organization Survey during other years.

Statistics on U.S. businesses from the Census Business Register are commonly used statistics for U.S. business activity. For example, these statistics are used in the National Income and Product Accounts (NIPA) and in County Business Patterns Statistics on U.S. businesses. Historically, Census has reported statistics on U.S. businesses separately for non-employer and employer businesses but recently the Census Bureau has been integrating the non-employer and employer components of the U.S.

business universe.[9] As part of that effort, research and database infrastructure work at the Center for Economic Studies has yielded early versions of an Integrated Longitudinal Business Database (ILBD). Davis et. al. (2005) use an early version of the ILBD to study the dynamics of U.S. Businesses with findings such as those discussed in section II (on Figure 1). In this regard, it is important to emphasize the respective sizes of the non-employer and employer portions of the U.S. business universe. Davis et. al. report that in the year 2000 there are roughly 15.5 non-employer firms and 5.5 employer firms in the U.S. While non-employer firms account for a large share of the total number of firms, they account for less than five percent of aggregate gross revenues from U.S. businesses. The latter finding might suggest that the non-employer businesses are relatively unimportant for measuring U.S. business activity but the findings we have already discussed for Figure 1 suggests that non-employers play an important role in the startup of employer businesses.

Because the unit of observation in the Business Register is a business rather than an individual, and because these two entities often differ, using universal business data to capture the universe of self-employed is not a straightforward task. We limit the scope of our analysis to those businesses where the relationship between the individual and the business is most clearly defined - sole proprietorships. Unfortunately, it is not feasible to include partnerships when defining the self-employed using the Business Register because very few partnerships in the Business Register have available a personal identifier for the business owners that may link, via the PIK, to the CPS and the DER. In the next section, we discuss the scope differences across datasets in more detail.

The source of sole proprietor data in the Census Bureau's Business Register is IRS Schedule C. Individuals filing IRS form 1040 use this schedule to report income earnings or losses from a business operated or a profession practiced as a sole proprietor. As stated in the instructions to schedule C, and re-stated at the top of Exhibit 3 in the appendix, "An activity qualifies as a business if your primary purpose for engaging in the activity is for income or profit and you are involved in the activity with continuity and regularity. For example, a sporadic activity or a hobby does not qualify as a business." The clear intent is to measure self-employment earnings rather than income from a hobby.

---

[9] The terms "non-employer" and "employer" are internal Census Bureau terms that warrant definition. According to http://www.census.gov/epcd/nonemployer/view/faq_nemp.html, "What is a non-employer business?" -- a non-employer business reported in official statistics is one that has no paid employees, and has annual business receipts of $1,000 or more ($1 or more in the construction industries.) Furthermore, "Are the non-employers the same as self-employed individuals?" – most non-employers are self-employed, although some partnerships and small corporations are included as well. Only partnerships and corporations with no paid employees are included in the non-employer statistics. Non-employers do NOT include about a million self-employed business owners that have paid employees and, therefore, are classified as employer businesses. Many non-employer businesses are part-time ventures and an individual might operate more than one. Note that we do not impose the $1000 positive receipts restriction that is used in official counts of non-employer businesses.

For purposes of generating statistics on gross revenue for U.S. businesses, the Census Bureau obtains the net receipts field from Schedule C. Net receipts are defined as gross receipts less the cost of the goods sold. However, self-employment earnings are measured in the CPS and the DER as net profits taking into account all business expenses. For the latter, there is less information available in the Census Business Register. In economic census years (years ending in "2" and "7"), a measure of non-agricultural self-employment earnings is available for non-employer sole proprietors, and this information in principle originates from the same source as self-employment earnings in the DER – line 2 of Schedule SE (net profit or loss from non-farm self-employment), which is also line 31 of Schedule C. In all other years, the only information available to measure earnings for non-employer sole-proprietorships is administrative current year net (of costs of good sold) receipts. For employer sole proprietors, Census has available from Schedule C as well as from form 941 (payroll tax forms) both receipts and payroll information. The latter two items yield a crude form of net earnings that can be used for employer sole-proprietors.

Basic descriptive statistics about sole proprietors (both employer and non-employer) and self-employment earnings for non-employers are given in Table 3. The sample in column 1 is the universe of sole proprietors filing Schedule C. The sample in column 2 is comprised of sole proprietors with a PIK match to the CPS.[10] In 1997, approximately 14.4 million sole proprietors are in the U.S. For the matched CPS-BR file, which is based upon the 74 percent of persons in the March CPS with a PIK, the number of sole proprietors is just over 12 million (on a weighted basis). Of the sole proprietors, about 13.4 million are non-employer sole proprietors and in the matched file roughly 9.6 million non-employer sole proprietors exist.

The mean and various distributional statistics for self-employment earnings from the BR non-employer sole proprietors are given in the bottom of Table 3. The mean and most of the percentiles are quite similar to the bottom panel of Table 2, which outlines the distribution of earnings from all self-employment jobs in the DER. This pattern is reassuring since the source information for the bottom panels of Table 2 and 3 are in principle the same. The 90th percentile is higher in the bottom panel of Table 2 relative to Table 3. This pattern also makes sense as the DER includes sole proprietors and partnerships with employees while the bottom panel of Table 3 is based only on the non-employer portion of sole proprietors. Employer businesses probably have higher net profits than non-employer businesses.

IIID. Scope Differences Across Datasets

The CPS should include all self-employed individuals including those who are incorporated and those with negative net earnings. As we have noted, we have excluded those CPS employed (whether self-employed or not) with non-positive total earnings from all jobs. We have also excluded incorporated self-employed in the CPS. We have

[10] As seen in Exhibit 3 of the appendix, the individual reports his or her SSN at the top of Schedule C. This SSN, converted to a PIK to protect the identity of the person, allows linkage of the Schedule C information to the March CPS and to the DER.

made these restrictions to keep the scope of the self-employed similar across datasets. In the DER, the positive net earnings threshold implies that negative net earnings workers are not included, and records of incorporated self-employed individuals are not covered by the DER. In the BR, there is a positive receipts requirement that should presumably hold even for negative net earnings workers. However, in the BR we have restricted our analysis to sole proprietors to avoid matching CPS incorporated self-employed workers to the BR.

Figure 3 illustrates the differences in scope across the two administrative datasets. The overlap between the BR sole proprietors and the DER self-employed are those sole proprietors with greater than $433 in net earnings. The DER will include partnerships not included in the BR sole proprietor universe, and the BR will include sole proprietors with net earnings below the DER threshold. The matched CPS-DER-BR data will allow us to quantify the respective sizes of the three areas in Figure 3 for all matched cases, which we show in section IV.D below.

III.E Employment in the March CPS and the DER

Before we begin our analysis of self-employment, we first analyze how the March CPS and the DER compare on employment status for calendar year 1997. Table 4 is a two-by-two matrix of employment status for the 137 million (weighted) individuals in both datasets. Employment in the March CPS is from the survey questions, and employment in the DER is based upon the W-2 forms filed for each individual by their employers, together with the Schedule SE filed by individuals as part of their 1040 form. An individual in the merged CPS – DER dataset is classified as not employed in the DER if no positive earnings record exists for that individual. The largest populated cell in the two-by-two matrix consists of 100 million individuals who are recorded as being employed by both datasets. These individuals represent 95.2 percent of employment in the DER and 90.9 percent of employment in the March CPS.

Table 4 illustrates that the March CPS and the DER do not always agree on employment status. The 10 million individuals who are recorded as working in the March CPS but who do not have a DER earnings record represent 9.0 percent of the CPS employed. Similarly, the 5 million individuals who are not employed in the March CPS but have a DER earnings record represent 4.8 percent of the DER employed.

We are not the first to find that the March CPS survey and the DER administrative data do not always agree on whether an individual was employed during a calendar year. Previous analysis of employment from a matched March CPS - DER dataset can be found in Roemer (2002). Roemer speculates that two possible explanations for employment in the March CPS but not the DER are first, the underground economy, where employers may fail to report an employee's wages to the Social Security Administration, and second, known reporting problems in the DER which cause under-coverage of workers in industries such as private households, agriculture,

and construction.[11] Roemer's tabulations (using data on an individual's longest non-self-employment job) show that childcare workers in private households, janitors and cleaners, farm workers, and private household cleaners and servants are the four most common occupations in the March CPS without matching DER records. One possible reason for the other off-diagonal, where an individual is employed in the DER administrative records but not in the March CPS, might be an individual forgetting periods of marginal employment from the previous year (which might be strongly influenced by self/proxy reporting in the CPS).

We believe that the misclassification rates on employment status between the March CPS and the DER – 9.0 percent of the March CPS employed do not have a DER earnings record, and 4.8 percent of the DER earnings record are recorded as not employed in the March CPS – set a baseline against which we can examine the misclassification of self-employment across the two datasets. Moreover, the discrepancy between CPS and DER employment status implies that in comparing self-employment rates and the mismatch in self-employment at the micro level, there are two considerations. First, do the CPS and DER agree on work status? Second, conditional on being employed, do the CPS and DER agree on the classification of self-employment vs. wage and salary positions? While we believe both of these questions are relevant, for the remainder of the analysis we focus on the 110 million workers who are classified as employed in the CPS. This permits us to focus on the second question and reduces the number of cases we consider considerably. However, given that the first question is quite relevant, in what follows we occasionally remark upon the sensitivity of the results to focusing on the CPS employed workers.

IIIF. Time Series Statistics

Because the Business Register provides information on self-employment earnings from Schedule SE only in Economic Census years, much of our analysis focuses on the year 1997. However, given the motivation and discussion in section II, it is clear that there is considerable interest in exploring how the administrative and household data compare in terms of time series fluctuations in self-employment.

Figure 4 presents self-employment rates for CPS employed workers using two alternative classifications of self-employed. Specifically, workers are classified as self-employed either on the basis of their responses to the CPS questionnaire or on the basis of their reported SE earnings in the DER. The top panel of Figure 4 shows that the self-employment rates for main jobs are similar in magnitude for CPS and DER classifications of self-employment status. However, switching occurs over time in which sources indicates higher self-employment rates. In the year including the recession, 2001, the CPS yields a decline in self-employment rates while the DER yields a mild increase. The differences in magnitudes at these high frequencies is as large as half of a percent which, as Figure 2 makes clear, is great enough to influence discrepancies between household and establishment data.

[11] A third explanation noted by Roemer is that about one percent of earnings records sent to the SSA can not be verified and posted to a specific individual's detailed earnings record.

The lower panel of Figure 4 shows the time series patterns for all jobs for CPS self-employed and DER self-employed conditional on being employed in the CPS. The DER self-employment rate is uniformly higher than the CPS self-employment rate. Moreover, the CPS self-employment rates exhibit a pronounced decline from 1997 to 2001.

To conclude this section, we provide a speculative discussion of the relationship between Figures 2 and 4. Recall from Figure 2 that in the period from 1997 through 2001, the CPS-adjusted employment (the largest two adjustments are for self-employment and for multiple job holders) statistic is lower than the CES payroll employment statistic. The gap exceeds one percent of employment in some years. Figure 4 shows that the disparity between the DER self-employment rates and CPS self-employment rates for main jobs is as much as half of a percent and is between one-half and two percent for all jobs. The magnitude of this gap is roughly comparable to the size of the gap in Figure 2 in at least 2001. However, the gap in Figure 4 appears to be the opposite of that in Figure 2. That is, in Figure 4 the DER-derived self-employment rate is higher than of the CPS. For purposes of discussion, if we were to assume that the DER rates were the truth, then too few of the employed CPS workers would be identified as being self-employed using CPS self-employment information. As such, the adjustment to the CPS for self-employment seen in Figure 2 should have been even larger in 2001.

IV. Microdata Comparisons of the Mismatch in Self-Employment Status

In the previous section, we discussed the aggregate levels of self-employment from three datasets as well as the scope differences across the datasets. In this section, we link the microdata across the three datasets and examine the propensity of a given individual to be classified as self-employed across sources. Our ultimate goal (not yet realized in this first draft) is to discern patterns in the misclassifications across datasets that will inform us about the quality of the self-employment data used by researchers and policy-makers.

IVA. Self-Employment in the March CPS and the DER – Basic Results

Tables 5a and 5b present a two-by-two matrix of self-employment status in the March CPS and the DER administrative records, conditional on being employed in the March CPS.[12] Table 5a presents the matrix for the "main job" definition of self-employment, and Table 5b presents the matrix for the "all jobs" definition of self-employment.

[12] As we have noted in section III, we could have presented Table 5 conditional on being employed in the DER (or conditional on being employed in both the March CPS and the DER). Since employment status is not consistent across the two datasets, the matrix of self-employment status across the two datasets will vary depending upon the sample. We have examined the analogous Table 5 conditional on DER employed and while the details differ, the qualitative patterns are quite similar.

In Table 5a, we see that 7 million employed individuals in the March CPS indicate they are self-employed in their main jobs (this is the same as column 2 in Table 1). We also see that 7 million of these 110 million individuals are recorded as self-employed in their main jobs according to the DER. Thus for these CPS employed individuals, the self-employment rates for main jobs are very similar across the two sources: in both the March CPS and the DER administrative records the self-employment rate is 6.4 percent.[13]

Although the aggregate self-employment rate is similar in the two datasets for this common group of workers, the individuals classified as self-employed in the two datasets are quite different. This is immediately evident by looking at the off-diagonals of Table 5a. 3 million individuals are self-employed in their main jobs in DER but not self-employed in the March CPS at their main job (and thus are classified as wage and salary workers in the CPS), and 3 million individuals are self-employed in their main jobs in the March CPS but not in the DER. Expressed as proportions, this level of misclassification is striking: 42.9 percent of the main job self-employed in the March CPS are not main job self-employed in the DER, and 42.9 percent of the main job self-employed in the DER are not main job self-employed in the March CPS.

Table 5b reports analogous statistics for all jobs. Here again the main finding is that while the rate of aggregate self-employment is similar across the CPS and the DER, the rate of mismatch is very high. Indeed, the mismatch rates (the off-diagonal percentages) in Table 5b are uniformly higher than the mismatch rates in Table 5a.

Roemer (2002) examined individuals who are not self-employed in the March CPS, but who have records of self-employment activity in the DER. He notes that the most commonly reported occupations of these individuals are real estate sales, barbers, fishers, social scientists not elsewhere classified, dentists, and clergy. Perhaps several of these occupations should not be considered misclassified – real estate agents earning commissions and clergy may have special treatment in the Social Security system and in the tax laws. Turning our focus to the other off-diagonal, where individuals are self-employed in the March CPS but not self-employed in the DER, one possible explanation for this discrepancy might be the $433 earnings threshold necessary to be considered self-employed in the DER. For main jobs this is unlikely, since in Tables 1 and 2 it is clear that for main jobs most CPS self-employment income is well above this threshold. Furthermore, self/proxy reporting in the March CPS might induce measurement error in self-reporting classifications. Later in this section, we conduct a regression analysis to help us identify the characteristics of individuals and jobs with different classifications of self-employment status across the three data sources.

IVB. Self-Employment in the March CPS and the Census Business Register

[13] This DER self-employment rate of 6.4 percent is lower than the 7.2 percent reported in Table 2 for main jobs. This is because the sample in Table 5b is restricted to CPS employed, and the group of individuals who are not employed in the March CPS but are employed in the DER (the 5 million persons, as reported in Table 4) have a higher self-employment rate (13.6 percent) than in the entire DER sample.

Table 6 presents a two-by-two matrix of self-employment status in the March CPS and the Census Bureau's Business Register (BR), conditional on being employed in the March CPS. Unlike the March CPS/DER comparison in Table 5b, the levels of self-employment are not similar in the March CPS and the BR. The self-employment rate in the BR is 11.6 percent, which is three percentage points higher than the self-employment rate of 8.6 percent in the March CPS. But similar to the findings from the March CPS – DER comparison, we find substantial amounts of misclassification of self-employment when comparing the microdata from the March CPS and the BR.

In Table 6, we see that 7.5 million individuals are self-employed in the BR but not self-employed in the March CPS, and 4.25 million individuals are self-employed in the March CPS but not in the BR. Again, this level of misclassification is striking: 44.7 percent of the self-employed in the March CPS are not self-employed in the BR, and 58.8 percent of the self-employed in the BR are not self-employed in the March CPS.

IVD. Self-Employment in the DER and the Census Business Register

To complete our analysis across the three datasets, Table 7 presents a two-by-two matrix of self-employment status in the DER administrative records and the Census Bureau's Business Register (BR), conditional on being employed in the March CPS. This matrix quantifies the various regions in Figure 3: 8 million individuals are classified as self-employed in both the DER and the BR. This large amount of overlap should not be surprising. Thinking back to the data description in the previous section (and in particular Exhibits 2 and 3 in the appendix), individuals who are sole-proprietors with more than $433 in self-employment income should in theory appear in both the DER and the BR.

Curiously, a large amount of individuals are classified as self-employed in either the DER or the BR, but not in both (conditional on being employed in the March CPS). Looking at the off-diagonals in Table 7, we find 5 million persons who are self-employed in the BR but not in the DER. According to Figure 3, these may be sole-proprietors with less than $433 in self-employment earnings. Looking at the other off-diagonal cell in Table 7, we find 2 million persons who are self-employed in the DER but not in the BR. According to Figure 3, these may be individuals in partnerships with more than $433 in self-employment earnings. In what follows, we explore the factors that impact the degree of agreement across self-employment status. One factor of clear interest in this context is self-employment earnings.

IVE. Exploring Misclassifications in Self-Employment Status Across Datasets – Regression Results

Tables 5, 6, and 7 show substantial disagreement of self-employment status across the three datasets (the March CPS, the DER administrative records, and the Census Bureau's Business Register). We now explore who the individuals are who are classified as self-employed in one dataset but not in the other, and attempt to determine whether any individual or job characteristics are systematically related to misclassification. In

Table 9, we present the results of probit regressions exploring the factors impacting the likelihood of agreement in self-employment status for main jobs. In these exercises, we examine type I and type II errors separately and in this way permit the impact of various worker and job characteristics on the probability of agreement to differ depending upon the nature of the disagreement. For example, in Table 5a, 57 percent of the workers who are identified as self-employed in their main jobs in the CPS are also identified as having their main job as self-employed in DER. Similarly, 57 percent of the workers who are identified as self-employed in their main jobs in the DER are also identified as having their main job as self-employed in the CPS. That is, we can consider separately the likelihood that a CPS self-employed worker is also identified as a DER self-employed worker (Model 1 in Table 8) and the likelihood that a DER self-employed worker is also identified as a CPS self-employed worker (Model 2 in Table 9).

In considering these models, we include the worker's age, education, gender, foreign born status, and race (a non-white indicator variable). We have conjectured that smaller business activities as well as contractual employment arrangements may both contribute to misclassifications across data sources in self-employment status. For this reason, we also include controls for industry and occupation as well as categories of self-employment earnings where the self-employment earnings measure used depends on the model (and thus the underlying sample at risk). For the probit using CPS main job self-employed workers as the sample (Model 1), the self-employment earnings measure used is the CPS main job self-employment earnings. For the probit using DER main job self-employed workers as the sample, the self-employment earnings measure used is the DER main job self-employment earnings. In estimating these models, we consider estimation with and without industry (1-digit) and occupation (2-digit) controls.

Summary statistics for the Model 1 and Model 2 samples are reported separately in Table 8. Although the age and racial distributions in these two samples are quite similar, the sample of DER main job self-employed (Model 2 sample) is somewhat more highly educated, has a higher percentage of males and foreign born workers and has a substantially higher share of workers earnings between $1,000 and $10,000 (about 40 percent versus 20 percent among CPS main job self-employed). This difference in the distribution of self-employment earnings across samples may reflect definitional differences in defining "main job." The main job in the CPS is defined as the longest held job during the year and the main job in the DER sample is defined as the job at which the worker received the highest earnings during the year.

Given that the focus is on factors that are associated with mismatch, a natural variable to explore is whether the individual is the respondent or another member of the family. Unfortunately, in 1997 the quality of this variable is poor so we do not use this variable. However, in what follows, we consider the robustness of our findings to other years for selected models (1998 and 1999) where the self-proxy variable is available and reliable. The limitation of the other years is that data from the BR register are less comparable as it is only in Economic Census years that the same self-employment income variable is available on the BR.

For Model 1, looking at the education variables, we find that the probability of agreement on self-employment status is largest for high school graduates and individuals with more than college degrees, with high school dropouts (the omitted category) having the lowest probability of agreement. No discernable trend by age is apparent, but there is considerable variation across age groups: the highest likelihood of agreement is for those who are between 35 and 44 years of age, with 25-34 year-olds (the omitted category) having the least likelihood of agreement. We find that the likelihood of a match in self-employment classification is higher for males, but this is only marginally statistically significant without industry/occupation controls and is insignificant when these controls are included in the specification. The likelihood of agreement on self-employment status is higher for foreign born workers, while non-whites have a significantly lower likelihood of agreement. For the most part, these patterns are robust to the inclusion of industry and occupation controls. In Model 2, we find less of a role for education, a substantially higher probability of agreement for workers that are above 35, a lower probability of matches for non-whites, and the probability of agreement on self-employment status to be much lower for foreign born workers. The patterns for Model 2 are more sensitive to the inclusion of industry and occupation controls.

The most robust pattern in Models 1 and 2 is the sensitivity to self-employment earnings. The greatest probability of disagreement in self-employment status is for workers with very low self-employment earnings. This finding is not surprising for two reasons: first, we have speculated that low self-employment earnings may be suggestive of either marginal employment or a "hobby business," and second, the earnings threshold in the DER will result in some CPS self-employed workers with low self-employment earnings not being classified as self-employed in the DER. However, it is important to note that even for high earnings workers, the likelihood of agreement in self-employment status is far from one. Figure 5 shows the predicted probabilities for 35-44 year-old males who are native born and are the CPS respondent (using models without industry and occupation controls). These predicted probabilities are shown for high school graduates and college graduates. Even for the highest earners, the predicted probabilities are less than 0.8. When one controls for industry and occupation, there is some variation both across models and across occupations. Figure 6 shows predicted probabilities for 35-44 year-old males who are native born and are the CPS respondent, for college educated accountants in the service industry as well as high school educated personal service workers in the service industry. For accountants, Model 2 yields higher probabilities than Model 1. In this case, it appears that accountants who are main job self-employed in the DER have a greater likelihood of also being self-employed at their main job in the CPS rather than the other way around. The differences between Models 1 and 2 are less pronounced for workers in the personal service occupations but the same pattern holds.

There are some stark differences in the results for Models 1 and 2. Primarily, we find that coefficients on controls for industry and occupation, while sometimes significant in Model 1, vary more in Model 2. This suggests that job characteristics play a much more prominent role in determining who among the DER self-employed will identify themselves as self-employed in the CPS. In contrast, CPS self-employed respondents

who are also self-employed in the DER do not differ by job traits (as proxied for by industry and occupation dummies) from those who are not self-employed in the DER. Not formally presented with these results is our finding that the DER self-employed who report themselves as working within the occupation of freight transport on the CPS questionnaire are associated with the highest negative impact on the probability of CPS self-employment. Note that this is an occupation characterized by a high percentage of contract workers. This finding is consistent with the conjecture that these types of workers may be less likely to identify themselves in household-based surveys as self-employed.

The impact of being foreign born on the probability of agreement also varies notably between Models 1 and 2. We find that for workers whose main job using the CPS is self-employment, being foreign born yields a somewhat greater likelihood of being identified as self-employed in a main job in the DER. However, for workers whose main job using the DER is self-employment, we find that being foreign born yields a much lower likelihood of being identified as self-employed in a main job in the CPS (the marginal effect when controlling for industry and occupation is about –12 percent). Note that in contrast, non-whites have a lower probability of agreement in a roughly symmetric way for both CPS and DER self-employed workers.

One feature of the differences between models 1 and 2 is that some of the patterns are not symmetric including education, age and foreign born status. The asymmetry suggests that it is not classical measurement error with heteroskedasticity across demographic groups driving these patterns. That is, for example, suppose the information used to define self-employment status were inherently noisier for foreign born workers. Then we would expect a lower likelihood of agreement by foreign born status in both Models 1 and 2. Given that we do find a lower probability of agreement for non-whites in both Models 1 and 2, this pattern might be driven by greater measurement error in the variables used to identify self-employment status for non-white individuals.

The asymmetric patterns suggest that there are some inherent differences in the determinants of self-employment across administrative and survey data. These asymmetric patterns deserve further exploration as they may reflect interesting differences in the type of self-employment that is identified across administrative and survey data as well as across demographic groups. For example, consider self-employed workers who are independent contractors but who work regularly for the same business – that is, self-employed workers where self-employment primarily reflects the form of the employer-employee relationship, rather than an independent business entity. It may be that such independent contractors are more likely to be identified as self-employed on the DER (since this will reflect tax filings) than on the CPS (where such a worker might be more likely to identify the main job as being employed by the business at which they regularly contract). Pushing harder on this example, if such independent contractors are more likely to be foreign born (other things equal) this could help explain the observed patterns.

When we examine the factors impacting the likelihood of agreement on self-employment status using all jobs, several different "models" are possible in the various combinations of comparing self-employment status in the CPS and DER, in the CPS and BR, and in the DER and BR. The results from these models for the 1997 sample are reported in Table 10 (where all reported results control for industry and occupation effects).[14] Models 1 and 2 in Table 10 relate to Table 5b, Models 3 and 4 relate to Table 6, and Models 6 and 7 relate to Table 7. While the details differ, one of the main messages from Table 9 also holds in this case. That is, the likelihood of agreement increases strongly with self-employment earnings. We also draw attention to Models 6 and 8. In Model 6, we find that for workers who have some self-employment earnings in DER, the probability of the worker being self-employed in the BR (defined as having positive BR net receipts: gross receipts reduced by the cost of goods sold) is very high (about 0.86 at the mean of the X variables, and 0.80 in Table 7). Moreover, we find in Model 8 that for workers who have some self-employment earnings in the BR (defined as positive SE earnings, from in principle the same IRS Schedule SE underlying the DER), the probability of being self-employed in the DER is almost one. Thus, it is reassuring that the two administrative datasets that seemingly draw upon the same source information agree on self-employment status. Also, on this last point, for Model 8 the likelihood of agreement is not related to the level of self-employment earnings. This makes sense as the two earnings concepts in BR and DER are in principle from the same source, and there is almost universal agreement for this subsample.

We also note that for Model 6, the role of earnings has a modest but different effect than in other models. Model 6 estimates the probability of being self-employed in the BR for the sample of self-employed from the DER (the mean of this dependent variable is 0.80 as seen in the appropriate row of Table 7). The estimates for Model 6 show that worker characteristics have little significance in explaining the agreement of self-employment status across the two datasets, and in particular the coefficients on the self-employment earnings variables are much different than in the other columns of the table, with the highest earnings category having a negative coefficient. We believe the explanation is straightforward: recall from Figure 3 that the DER and the BR have a substantial overlap in self-employment status, and in principle the self-employed in the DER who are not self-employed in the BR are individuals in business partnerships. If partners earn more than sole-proprietorships, and part of this earnings differential is not picked up by occupation, then the negative earnings coefficients in Model 6 make sense.

[14] For the BR two different samples are used in Table 9 along with two different concepts of self-employment earnings. One sample is all matched BR sole proprietors with positive net receipts from Schedule C – this is the sample that underlies our analysis in Tables 6 and 7. The earnings concept here is net receipts (gross receipts minus cost of goods sold). The second (smaller) sample is all matched BR non-employer sole proprietors with positive Schedule SE net earnings (recall this concept is identical to earnings concept in DER and is a net profits concept). The careful reader of Table 9 might also note differences in sample sizes across models with seemingly the same underlying sample of individuals. For example, models 1 and models 3 are both based on the sample of CPS workers with any self-employment income but model 3 has more observations than model 1. The reason is that in the models in Table 9 industry and occupation controls are included and in some cases industry and occupation together yield perfect predictions of status so those observations are excluded from the estimation.

To evaluate whether these patterns hold up in other years, we supplement these 1997 findings by estimating these same "main job" Model 1 and Model 2 specifications using pooled 1998 and 1999 data on self-employment activity. Though not shown in the paper, the overall probabilities of mismatch for main job self-employment (the off-diagonal elements in tables 5a and 5b) are similar in all 3 years. In addition, the distributions of key explanatory variables are similar in all years as well. Data in these later years have the additional feature of permitting us to include a dummy variable indicating whether the individual is the CPS household respondent.[15] Main job estimates using pooled 1998 and 1999 data are presented in Table 11. For each model and year, we present estimates with a variable indicating that the individual is the CPS household respondent.

Being the CPS household respondent does not impact the probability of agreement in Model 1 and has only a very small positive impact in Model 2. Thus, those individuals who we find to be main job self-employed in the DER are more likely to be identified as main job self-employed in the CPS if they are the household respondent. Though not presented, estimates were generated with and without this "respondent" indicator variable included, and it is worth noting that even in Model 2, inclusion of the respondent indicator variable has virtually no impact on other coefficients. Thus, differences between the 1997 and pooled 1998 and 1999 estimates do not result simply from the inclusion of the "respondent" control in the later year specification.

Comparing the 1998 and 1999 results for Models 1 and 2 to those for 1997, we first note that the predicted probabilities of these models are quite similar to those for 1997. For example, the predicted probability of agreement (evaluated at variable means) using both 1997 and 1998/1999 samples for Model 2 are .62 with industry and occupation controls and .60 without. In addition, we find that the strongest patterns hold up across years. Most notably, coefficients on the controls for earnings category, being foreign born and being non-white are fairly similar in 1998/1999 to 1997. The asymmetric impact of the foreign born indicator in models 1 and 2 holds up as well. There are, however, a few differences. Although the lower probability that DER main job self-employed workers who are foreign born indicate that they are self-employed at their main job in the CPS (Model 2) does hold up, the positive impact of being foreign born for Model 1 loses significance. Additionally, the negative non-white impact diminishes in Model 2. Finally, those demographic traits that have a weaker link to the probability of self-employment agreement in 1997 (such as age, gender, and education) in many cases emerge more strongly in the pooled sample regressions.

We estimate models 1 and 2 for the "all jobs" specifications considered above using the pooled 1998/1999 samples as well. These results are presented in columns 3 and 6 of table 11. These estimates are quite similar to those obtained from the 1997 sample and demonstrate less variability across years than do the "main job" estimates.

V. Conclusions

[15] A clean measure of this variable is not available in the internal March Supplement containing self-employment information for 1997.

By linking administrative and survey data on self-employment, we have been able to evaluate the relationship between micro and aggregate self-employment rate patterns from alternative sources. Our primary (and preliminary) findings are as follows:

- The self-employment rates from household surveys and administrative data are roughly similar but the differences are non-trivial in magnitude and vary over time (Figure 4). While it is premature to judge whether these discrepancies are sufficient to help account for differences in household- and establishment-based employment statistics, it appears as if the direction of the disparity between the household and administrative data goes the "wrong" way. If anything, our results suggest that the March CPS may be under-counting self-employment.
- While the aggregate magnitudes are roughly similar, we find a large amount of disagreement at the micro level about self-employment status. Using main jobs, for example, we find that less than 60 percent of those identified as self-employed in the CPS are also self-employed in the administrative data (Table 5a).
- The most important factor that impacts the extent of agreement in self-employment status at the micro level turns out to be self-employment earnings (Table 8). Some of this appears to be due to income thresholds in at least some of the administrative data to be classified as self-employed. However, even for workers with earnings well above these thresholds the likelihood of agreement on self-employment status is far from one.
- Demographic factors such as age, education, race, and foreign born status also impact the likelihood of agreement in self-employment status. In some cases, there is a distinct asymmetry in the impact of these characteristics on whether a CPS self-employed worker is a DER self-employed worker and vice versa. For example, a foreign-born CPS worker whose main job is self-employment has a 36 percent higher chance of also being identified as a DER worker whose main job is self-employment. In contrast, a foreign-born DER worker whose main job is self-employment has a 12 percent lower chance of also being identified as a CPS worker whose main job is self-employment. We do not have a ready explanation for such stark asymmetries, but these findings suggest caution about the inferences drawn from studies about the self-employment status by demographic groups and perhaps especially about inferences about the volatility of different demographic groups in terms of their propensity to be self-employed. Given the degree of mismatch between household and administrative data for the same individuals, measurement or conceptual differences in self-employment appear to exist that vary by key demographic groups.
- Although the CPS may provide better coverage of certain types of more traditional self-employment activity, it is often asserted that changes in the economy have altered the nature of self-employment. More detailed probing into the nature of work activities in which respondents engage

may be required for the March CPS to maintain its ability to measure self-employment activity in the U.S.

- We have not yet done a comparison of self-employment earnings across datasets.

At this preliminary stage of this study, the comparison of administrative and survey data seems to indicate that no single source precisely measures entrepreneurs, as each is subject to its own coverage and measurement issues. However, we have gained insights into circumstances under which these administrative records may themselves be used as an independent source of data on self-employment and entrepreneurial activity. To acquire the best concept of national employment (and in particular self-employment rates), as well as to study the micro dynamics of self-employment and entrepreneurship, our study suggests that it is likely to be important to understand and incorporate the discrepancies between survey and administrative data.

References

Abowd, John and Martha Stinson, "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and Administrative Data" September 2003.

Bednarzik, Robert W. "The Role of Entrepreneurship in U.S. and European Job Growth," *Monthly Labor Review*, July 2000, pp. 3–16.

Blanchflower, David G. "Self-Employment in OECD Countries," *Labour Economics*, September 2000, pp. 471–506.

Blanchflower, David G. and Andrew J. Oswald, "What Makes an Entrepreneur?" *Journal of Labor Economics*, January 1998, pp. 26–60.

Blanchflower, David G. Andrew J. Oswald, and Alois Stulzer, "Latent Entrepreneurship across Nations," *European Economic Review*, May 2001, pp. 680–691.

Blau, David M. 1987. "A Time-Series Analysis of Self-Employment in the United States." *The Journal of Political Economy*, Vol. 95, No. 3, pp. 445-467.

Boden, Richard and Al Nucci. 1995. "Counting the Self-Employed from Two Perspectives: Household vs. Business Sample Data" CES Working Paper No. 95-11, August.

Bowler, Mary and Tereas L. Morisi. "Understanding the employment measures from the CPS and CES survey." *Monthly Labor Review*, February 2006, pp. 23-38.

Bregger, John E. "Measuring Self-Employment in the United States," *Monthly Labor Review*, January/February 1996, pp. 3–9.

Carrington, William J., Kristin McCue, and Brooks Pierce. 1996. "The Role of Employer/Employee Interactions in Labor Market Cycles: Evidence from the Self-Employed." *Journal of Labor Economics*, Vol. 14, No. 4, pp. 571-602.

Davis, Steven J., John Haltiwanger, Ron Jarmin, C.J. Krizan, Javier Miranda, Al Nucci and Kristin Sandusky. 2006. "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes", CES Working Paper No. 06-04, February.

Deming, William G. "Work at Home: Data from the CPS," *Monthly Labor Review*, February 1994, pp. 14–20.

Devine, Theresa "Characteristics of self-employed women in the United States," *Monthly Labor Review*, March 1994, pp. 20–34.

Devine, Theresa. 1994. "Changes in Wage-and-Salary Returns to Skill and the Recent Rise in Female Self-Employment." *The American Economic Review*, Vol. 84, No. 2, pp. 108-113.

Edwards, Linda N. and Elizabeth Field-Hendrey, "Home-Based Work and Women's Labor Force Decisions," *Journal of Labor Economics*, January 2002, pp. 170–200.

Fairlie, Robert W. *Ethnic and Racial Entrepreneurship: A Study of Historical and Contemporary Differences* (New York, Garland, 1996).

Fairlie, Robert. 2005." Entrepreneurship among Disadvantaged Groups: An Analysis of the Dynamics of Self-Employment by Gender, Race, and Education", forthcoming in Handbook of Entrepreneurship (eds. Simon C. Parker, Zoltan J. Acs, and David R. Audretsch), Kluwer Academic Publishers

Fairlie, Robert W. and Bruce D. Meyer. 1994. "The Ethnic and Racial Character of Self-Employment," NBER Working Paper No. 4791 (National Bureau of Economic Research).

Fairlie, Robert W. and Bruce D. Meyer. 2000. "Trends in Self-Employment among White and Black Men during the Twentieth Century." *The Journal of Human Resources*, Vol. 35, No. 4, pp. 643-669.

Fairlie, Robert W. and Bruce D. Meyer, "Ethnic and Racial Self-Employment." 44 Monthly Labor Review, July 2004.

Fuchs, Victor R. "Self-Employment and Labor Force Participation of Older Males," Journal of Human Resources, Summer 1982, 339–57.

Genda, Yuji and Ryo Kambayashi, "Declining Self-Employment in Japan," *Journal of the Japanese and International Economies*, March 2002, pp. 73–91.

Georgellis, Yannis and Howard J. Wall, "Who Are the Self-Employed?" *Federal Reserve Bank of St. Louis Review*, November/December 2000, pp. 15–23.

Gruber, Jonathan and James Poterba. 1994. "Tax Incentives and the Decision to Purchase Health Insurance: Evidence from the Self-Employed." *The Quarterly Journal of Economics*, Vol. 109, No. 3, pp. 701-733.

Haider, Steven and D. Loughran. 2001. "Elderly Labor Supply: Work or Play?" RAND DRU-2582, Santa Monica, CA.

Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *The Journal of Political Economy*, Vol. 108, No. 3, pp. 604-631.

Hipple, Steven. 2004. "Self-employment in the United States: an update." *Monthly Labor Review*, Vol. 127, No. 7, pp. 13-23.

Hundley, Greg. "Why Women Earn Less Than Men in Self-Employment," *Journal of Labor Research*, Fall 2001, pp. 817–29.

Kalton, G., and Flores-Cervantes, I., (2003) "Weighting Methods," Journal of Official Statistics, Vol. 19, No. 2, 81-97.

Manser, Marilyn E. and Garnett Picot, "The role of self-employment in U.S. and Canadian job growth, *Monthly Labor Review*, April 1999, pp. 10–25.

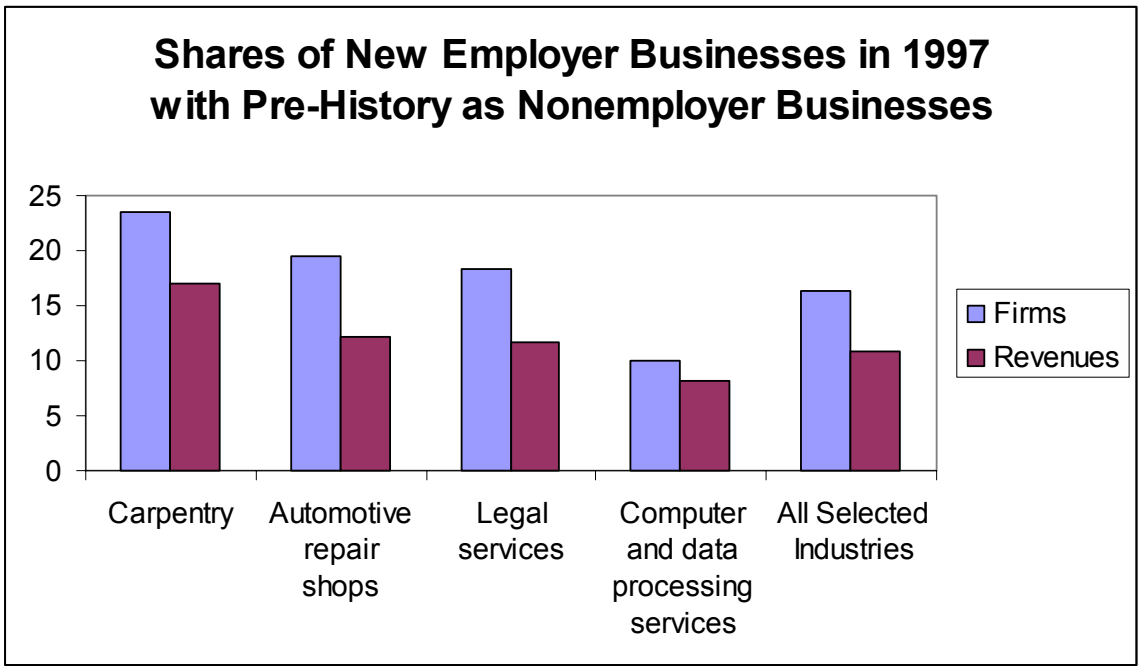
Roemer, Marc, "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation." LEHD Technical Paper #TP-2002-22, November 2002.

Quinn, Joseph. "Labor Force Participation Patterns of Older Self-Employed Workers," *Social Security Bulletin*, April 1980, pp. 17–28.

Zissimopoulos, Julie and Lynn A. Karoly, "Transitions to Self-Employment at Older Ages: The Role of Wealth, Health, Health Insurance, and Other Factors," paper presented at the *Annual Meetings of the Population Association of America*, April 2003.

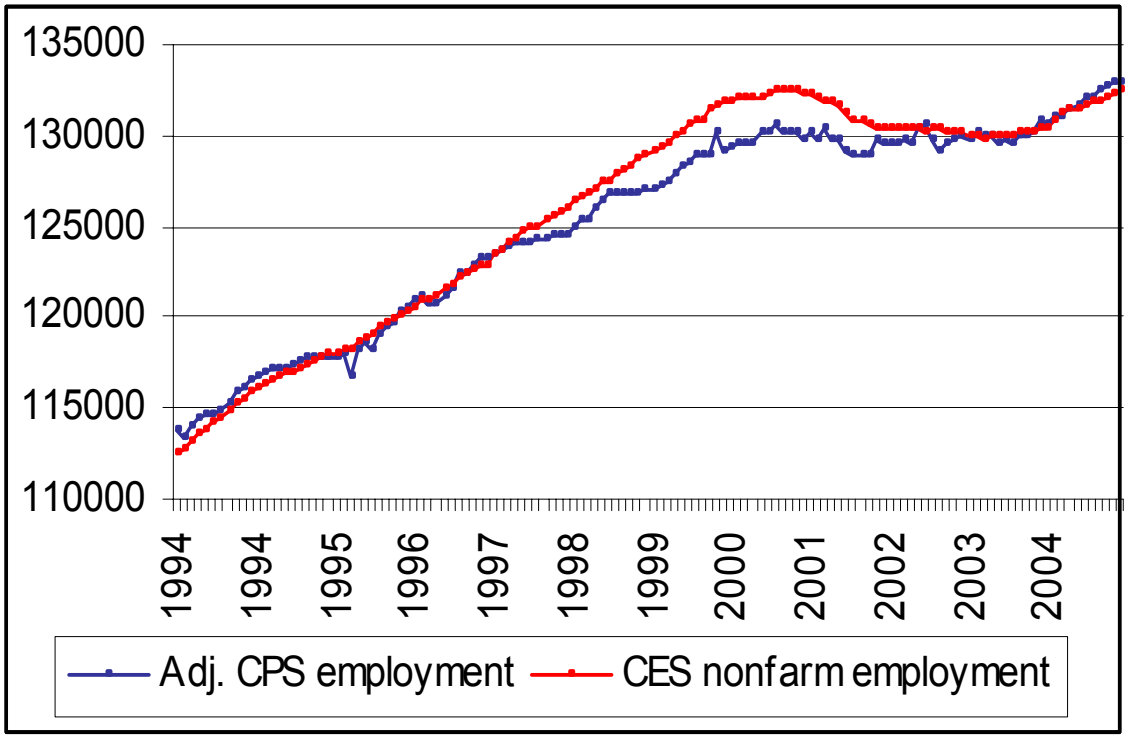
Wun L, Ezzati-Rice TM, Baskin R, Greenblatt J, Zodet M, Potter F, Diaz-Tena N, Touzani M. Using Propensity Scores to Adjust Weights to Compensate for Dwelling Unit Level Nonresponse in the Medical Expenditure Panel Survey. Agency for Healthcare Research and Quality Working Paper No. 04004, October 2004, http://www.ahrq.gov.

Figure 1



Source: Davis et. al. (2006)

Figure 2. Comparison of Employment Statistics from Household and Establishment Surveys



Source: BLS

Figure 3: Scope Differences between the DER and the BR

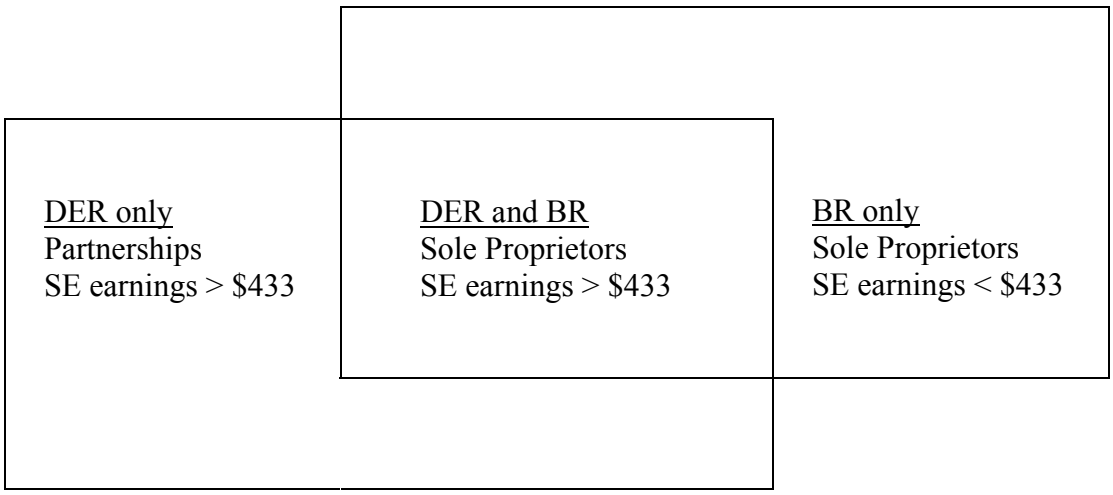


Figure 4a: Self-Employment in the March CPS and DER
Sample: Employed in the CPS
Using "Main Job" Definition of Self-Employment

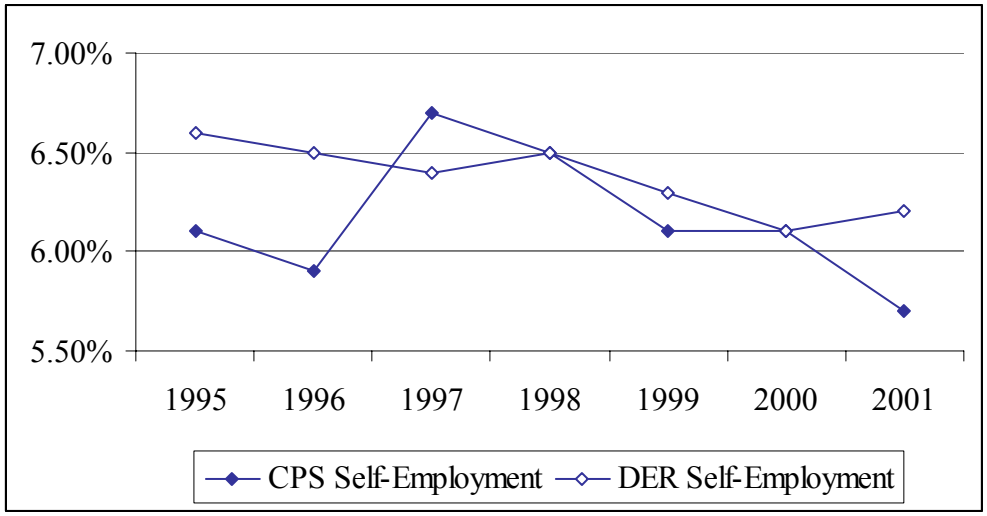


Figure 4b: Self-Employment in the March CPS and DER
Sample: Employed in the CPS
Using "All Jobs" Definition of Self-Employment

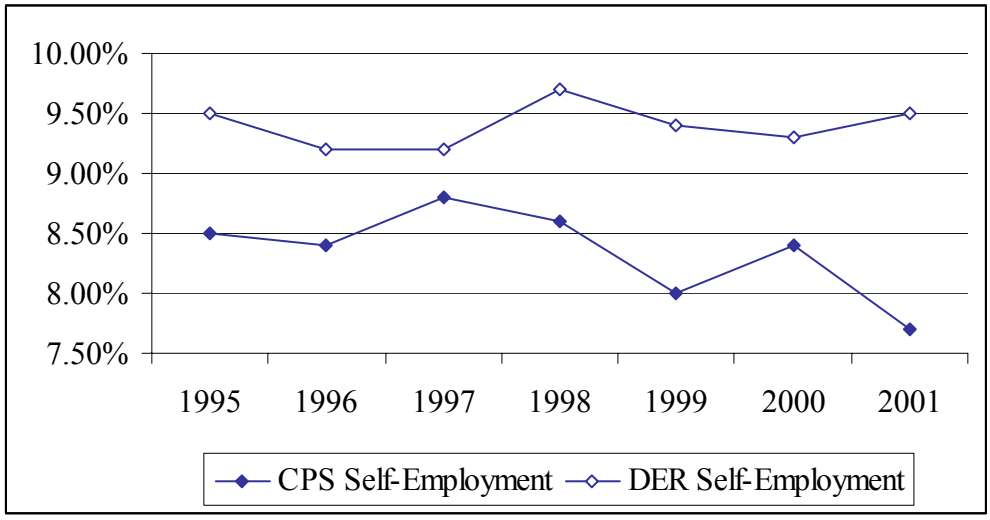


Figure 5

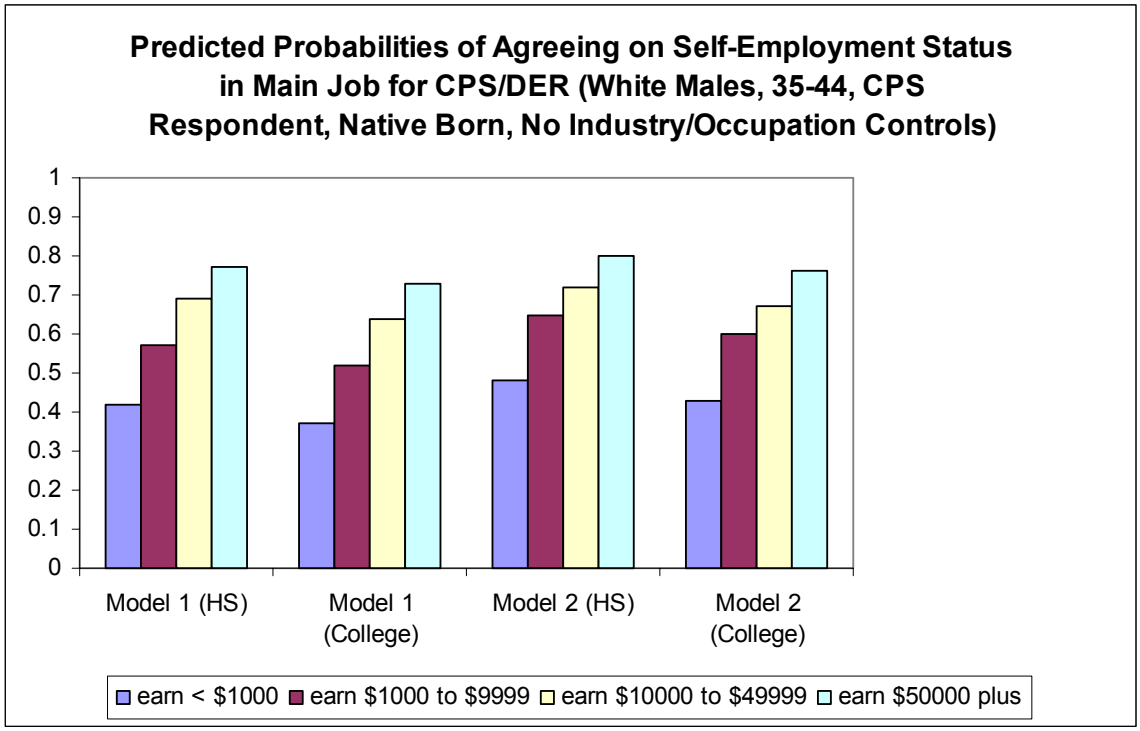

Predicted Probabilities of Agreeing on Self-Employment Status in Main Job for CPS/DER (White Males, 35-44, CPS Respondent, Native Born, No Industry/Occupation Controls)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Model 1 (HS)
Model 1 (College)
Model 2 (HS)
Model 2 (College)
earn < $1000
earn $1000 to $9999
earn $10000 to $49999
earn $50000 plus

Figure 6

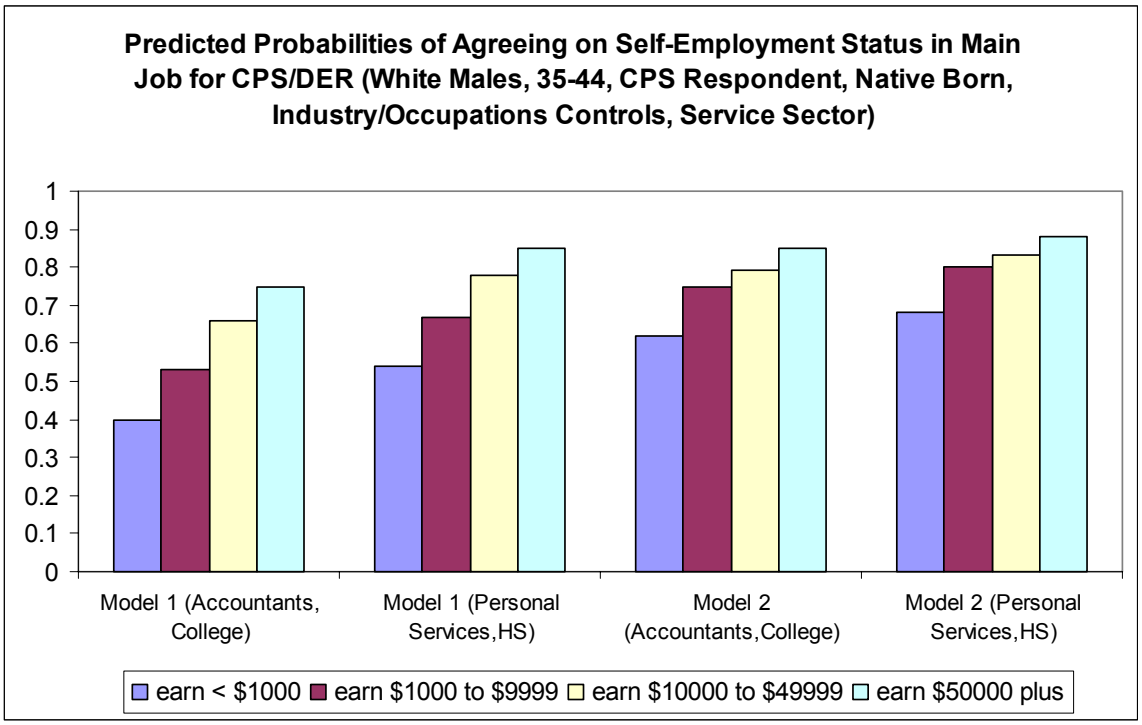

Predicted Probabilities of Agreeing on Self-Employment Status in Main Job for CPS/DER (White Males, 35-44, CPS Respondent, Native Born, Industry/Occupations Controls, Service Sector)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Model 1 (Accountants, College)
Model 1 (Personal Services,HS)
Model 2 (Accountants,College)
Model 2 (Personal Services,HS)
earn < $1000
earn $1000 to $9999
earn $10000 to $49999
earn $50000 plus

Table 1: Self-Employment in 1997 from the 1998 March CPS

	Full Sample	Full Sample With PIK
# Persons	137.1M	137.1M
# Employed Persons	110.1M	110.1M
# Self-Employed Persons, Main Job	7.5M	7.3M
Self-Employment Rate, Main Jobs	6.82%	6.61%
# Self-Employed Persons, All Jobs	9.6M	9.6M
Self-Employment Rate, All Jobs	8.72%	8.72%
Self-Employed Earnings, Main Job		
Mean	31,291	30,370
10^{th} Percentile	1,900	1,660
25^{th} Percentile	8,100	7,360
50^{th} Percentile	19,600	19,175
75^{th} Percentile	36,800	36,400
90^{th} Percentile	66,000	65,800
Self-Employed Earnings, All Jobs		
Mean	26,719	25,716
10^{th} Percentile	300	300
25^{th} Percentile	4,000	3,513
50^{th} Percentile	14,400	13,200
75^{th} Percentile	31,400	30,400
90^{th} Percentile	59,400	58,000
Unweighted Sample Size	66,397	49,877

All data are restricted to persons aged 25-65. We exclude any individual whose main job during the year was agriculture, forestries, or fisheries, or who was in the Armed Forces. Counts are reported in millions. Percentiles are approximate percentiles computed as the average of the five centiles around the listed percentile, rounded to the nearest $100. All data except the bottom row are weighted by the adjusted March CPS sample weight. PIK = Personal Identification Key.

Table 2: Self-Employment in the 1997 SSA Detailed Earnings Record

	Sample with Merge to CPS
# Employed Persons	105.1M
# Self-Employed Persons, Main Job	7.6M
Self-Employment Rate, All Jobs	7.23%
# Self-Employed Persons, All Jobs	10.8M
Self-Employment Rate, All Jobs	10.3%
Self-Employed Earnings, Main Job	
Mean	19,300
10th Percentile	1,600
25th Percentile	4,400
50th Percentile	10,700
75th Percentile	25,900
90th Percentile	56,300
Self-Employed Earnings, All Jobs	
Mean	15,800
10th Percentile	1000
25th Percentile	2,300
50th Percentile	6,900
75th Percentile	18,800
90th Percentile	47,000
Unweighted Sample Size	49,877

All data except the last row are weighted by the adjusted March CPS sample weight. Counts are reported in million. Percentiles are approximate percentiles computed as the average of the five centiles around the listed percentile, rounded to the nearest $100.

Table 3: Self-Employment in the 1997 Business Register Non-employer File

	Universe	Sample With Merge to CPS
# Sole Proprietors	14.5M	13.5M
# Non-employer Sole Proprietors	13.4M	12M
# Non-employer Sole Proprietors with non-missing self-employment income	9.1M	7.8M
Self-Employed Earnings		
Mean	15,800	15,500
10th Percentile	1,000	1000
25th Percentile	2,300	2,200
50th Percentile	6,400	6,300
75th Percentile	15,500	16,200
90th Percentile	34,700	37,200
Unweighted Sample Size		49,877

Data in the first column are unweighted. Data in column 2 are weighted by the adjusted March CPS sample weight except for the last row. Counts are reported in millions. Percentiles are approximate percentiles computed as the average of the five centiles around the listed percentile, rounded to the nearest $100.

Table 4: Employment in 1997, comparing the March CPS and the DER

	Not Employed in DER	Employed in DER	Row Total
Not Employed in CPS	22M	5M	27M
Percent of Total	16.1%	3.6%	
Row Percent	81.5%	18.5%	
Column Percent	68.8%	4.8%	19.7%
Employed in CPS	10M	100M	110M
Percent of Total	7.3%	73%	
Row Percent	9.1%	90.9%	
Column Percent	31.2%	95.2%	80.3%
Column Total	32M	105M	137M
Row Percent	23.4%	76.6%	

All data are weighted by the adjusted March CPS sample weight.

Table 5a: Self-Employment in 1997, comparing the March CPS and the DER
Sample: Employed in the CPS
Using "Main Job" Definition of Self-Employment

	Not Self-Employed in DER	Self-Employed in DER	Row Total
Not Self-Employed in CPS	100M	3M	103M
Percent of Total	90.9%	2.7%	
Row Percent	97.1%	2.9%	
Column Percent	97.1%	42.9%	93.6%
Self-Employed in CPS	3M	4M	7M
Percent of Total	2.7%	3.6%	
Row Percent	42.9%	57.1%	
Column Percent	2.9%	57.1%	6.4%
Column Total	103M	7M	110M
Row Percent	93.6%	6.4%	

All data are weighted by the adjusted March CPS sample weight.

Table 5b: Self-Employment in 1997, comparing the March CPS and the DER
Sample: Employed in the CPS
Using "All Jobs" Definition of Self-Employment

	Not Self-Employed in DER	Self-Employed in DER	Row Total
Not Self-Employed in CPS	95.5M	5M	100.5M
Percent of Total	86.8%	4.5%	
Row Percent	95%	5%	
Column Percent	95.5%	50%	91.4%
Self-Employed in CPS	4.5M	5M	9.5M
Percent of Total	4.2%	4.5%	
Row Percent	47.4%	52.6%	
Column Percent	4.5%	50%	8.6%
Column Total	100M	10M	110M
Row Percent	90.9%	9.1%	

All data are weighted by the adjusted March CPS sample weight.

Table 6: Self-Employment in 1997, comparing the March CPS and the BR
Sample: Employed in the CPS
Using "All Jobs" Definition of Self-Employment

	Not Self-Employed in BR	Self-Employed in BR	Row Total
Not Self-Employed in CPS	93M	7.5M	100.5M
Percent of Total	84.5%	6.8%	
Row Percent	92.5%	7.5%	
Column Percent	95.6%	58.8%	91.4%
Self-Employed in CPS	4.25M	5.25M	9.5M
Percent of Total	3.9%	4.8%	
Row Percent	44.7%	55.3%	
Column Percent	4.4%	41.2%	8.6%
Column Total	97.25M	12.75M	110M
Row Percent	88.4%	11.6%	

All data are weighted by the adjusted March CPS sample weight.

Table 7: Self-Employment in 1997, comparing the DER and the BR
Sample: Employed in the CPS
Using "All Jobs" Definition of Self-Employment

	Not Self-Employed in BR	Self-Employed in BR	Row Total
Not Self-Employed in DER	95M	5M	100M
Percent of Total	86.4%	4.5%	
Row Percent	95.0%	5.0%	
Column Percent	97.9%	38.5%	90.9%
Self-Employed in DER	2M	8M	10M
Percent of Total	1.8%	7.3%	
Row Percent	2.0%	80.0%	
Column Percent	2.1%	61.5%	9.1%
Column Total	97M	13M	110M
Row Percent	88.2%	11.8%	

All data are weighted by the adjusted March CPS sample weight.

Table 8: 1997 Main Job Regression Sample Summary Statistics (Means of Explanatory Variables)

	Model 1: CPS Self-employed Main Job	Model 2: DER Self-employed Main Job
Number of Observations	2,793	2,696
age 25 to 34	0.186	0.203
age 35 to 44	0.339	0.330
age 45 to 54	0.287	0.284
age 55 to 65	0.187	0.182
Less than High School	0.112	0.105
High School Graduate	0.311	0.315
Some College	0.272	0.228
College Graduate	0.190	0.195
Post College	0.115	0.158
Male	0.605	0.644
Foreign Born	0.111	0.147
Non-white	0.083	0.088
Self-Emp Earnings < $1,000.	0.078	0.042
Self Emp Earnings $1000 to $9,999	0.217	0.411
Self Emp Earnings $10,000 to $49,000	0.535	0.421
Self-Emp Earnings $50,000 Plus	0.170	0.125

Note: All Variables are categorical so the reported means can be interpreted as the fraction of individuals with the reported characteristic

Table 9: Marginal Effects from Probit Analysis on Agreement on Self-Employment Status for Main Jobs (CPS vs. DER Self-Employed Workers)

	Model 1		Model 2	
Sample	CPS Main Job SE		DER Main Job SE	
Number of Observations, unweighted	2,789		2,592	
Dependent Variable	DER Main Job SE		CPS Main Job SE	
Self-Employment Earnings Measure	SE Earn CPS Main Job		SE Earn DER Main Job	
Industry and Occupation Controls Included	Yes	No	Yes	No
Age 35 to 44	0.072	0.062	0.148	0.123
	(0.031)*	(0.031)*	(0.031)**	(0.029)**
Age 45 to 54	0.042	0.028	0.137	0.087
	(0.033)	(0.032)	(0.031)**	(0.030)**
Age 55 to 65	0.065	0.043	0.166	0.107
	(0.036)	(0.036)	(0.034)**	(0.033)**
High School Graduate	0.115	0.100	0.043	0.048
	(0.038)**	(0.037)**	(0.040)	(0.038)
Some College	0.020	-0.007	0.058	0.032
	(0.040)	(0.039)	(0.043)	(0.040)
College Graduate	0.105	0.048	0.091	-0.006
	(0.043)*	(0.040)	(0.046)*	(0.042)
Post College	0.144	0.135	-0.080	-0.117
	(0.053)**	(0.044)**	(0.061)	(0.046)**
Male	0.049	0.051	-0.048	-0.008
	(0.030)	(0.024)*	(0.031)	(0.023)
Foreign Born	0.100	0.072	-0.111	-0.094
	(0.036)**	(0.035)*	(0.036)**	(0.033)**
Non-white	-0.158	-0.154	-0.159	-0.115
	(0.043)**	(0.042)**	(0.047)**	(0.042)**
Self Emp Earnings $1000 to $9,999	0.126	0.141	0.142	0.163
	(0.045)**	(0.044)**	(0.057)*	(0.053)**
Self Emp Earnings $10,000 to $49,000	0.264	0.269	0.181	0.237
	(0.043)**	(0.042)**	(0.058)**	(0.052)**
Self Emp Earnings $50,000 or above	0.322	0.321	0.243	0.297
	(0.038)**	(0.037)**	(0.051)**	(0.042)**
Predicted Probability at Mean X	0.57	0.57	0.62	0.60

Table 10: Marginal Effects from Probit Analysis on Agreement of Self-Employment Status for All Jobs

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Sample	CPS SE any	DER SE any	CPS SE Any	BR Sole Prop Receipts	BR SE earnings (Net Earnings)	DER SE Any	BR Sole Prop Receipts	BR SE Earnings (Net Earnings)
# Observations, unweighted	3,651	3,900	3,664	4,876	3,006	3,900	4,876	2,034
Dependent Variable	DER SE Any	CPS SE Any	BR Sole Prop Receipts	CPS SE Any	CPS SE Any	BR Sole Prop Receipts	DER SE Any	DER SE Any
SE Earnings Measure Used	CPS Total SE Earnings	DER Total SE Earnings	CPS Total SE Earnings	BR Sole Prop Receipts	BR SE Earnings	DER Total SE Earnings	BR Sole Prop Receipts	BR SE Earnings
Age 35 to 44	0.039	0.120	0.031	0.063	0.107	0.017	-0.042	-0.027
	(0.028)	(0.027)**	(0.027)	(0.025)*	(0.031)**	(0.018)	(0.024)	(0.012)*
Age 45 to 54	-0.001	0.106	0.048	0.076	0.113	0.030	-0.049	-0.013
	(0.029)	(0.028)**	(0.028)	(0.027)**	(0.032)**	(0.018)	(0.025)*	(0.010)
Age 55 to 65	0.024	0.146	0.118	0.119	0.141	0.061	-0.067	-0.016
	(0.034)	(0.032)**	(0.031)**	(0.031)**	(0.037)**	(0.018)**	(0.029)*	(0.015)
High School Graduate	0.083	0.035	0.083	-0.000	0.004	-0.000	-0.042	-0.008
	(0.037)*	(0.036)	(0.035)*	(0.033)	(0.042)	(0.026)	(0.037)	(0.020)
Some College	0.004	0.052	0.042	0.045	0.047	-0.034	-0.120	-0.058
	(0.038)	(0.039)	(0.037)	(0.036)	(0.045)	(0.028)	(0.039)**	(0.037)
College Graduate	0.102	0.117	0.103	0.059	0.114	-0.010	-0.097	-0.068
	(0.041)*	(0.041)**	(0.039)**	(0.038)	(0.048)*	(0.029)	(0.041)*	(0.040)
Post College	0.094	-0.018	0.102	0.007	-0.025	-0.031	-0.108	-0.141
	(0.049)	(0.050)	(0.048)*	(0.046)	(0.057)	(0.036)	(0.047)*	(0.065)*
Male	0.031	-0.038	0.141	-0.106	-0.048	-0.025	-0.284	-0.036
	(0.026)	(0.025)	(0.025)**	(0.024)**	(0.029)	(0.016)	(0.017)**	(0.007)**
Foreign Born	0.094	-0.076	0.041	-0.125	-0.124	0.042	0.061	0.001
	(0.034)**	(0.031)*	(0.032)	(0.026)**	(0.035)**	(0.020)*	(0.026)*	(0.009)
Non-white	-0.164	-0.145	-0.136	-0.098	-0.155	-0.025	-0.042	0.009
	(0.036)**	(0.036)**	(0.037)**	(0.031)**	(0.041)**	(0.028)	(0.031)	(0.008)
SE Earnings $1,000-$9,999	0.176	0.157	0.152	0.212	0.172	-0.028	0.293	0.011
	(0.033)**	(0.037)**	(0.031)**	(0.038)**	(0.041)**	(0.024)	(0.025)**	(0.008)
SE Earnings $10,000-$49,999	0.324	0.291	0.210	0.432	0.318	-0.021	0.418	0.012
	(0.030)**	(0.036)**	(0.030)**	(0.034)**	(0.041)**	(0.026)	(0.020)**	(0.008)
SE Earnings $50,000+	0.378	0.361	0.200	0.588	0.352	-0.126	0.460	0.015
	(0.028)**	(0.035)**	(0.035)**	(0.028)**	(0.042)**	(0.041)**	(0.015)**	(0.005)**
Predicted Probability at Mean X	0.53	0.51	0.59	0.39	0.49	0.86	0.69	0.98

Note: Controls for industry and occupation are included in all specifications. * significant at 5%; ** significant at 1%

Table 11: **Marginal Effects from Pooled 1998-1999 Probit Analysis on Agreement on Self-Employment Status (CPS vs. DER Self-Employed Workers)**

	Model 1			Model 2		
Sample	CPS Main Job SE		CPS All Job SE	DER Main Job SE		DER All Job SE
Dependent Variable	DER Main Job SE		DER All Job SE	CPS Main Job SE		CPS All Job SE
Self-Employment Earnings Measure	SE Earn CPS Main Job		SE Earn CPS All	SE Earn DER Main Job		SE Earn DER All
Industry and Occupation Controls Included	Yes	No	Yes	Yes	No	Yes
Number of Observations, un-weighted	4877	4887	6398	4527	4955	7359
Age 35 to 44	0.061	0.050	0.043	0.090	0.075	0.070
	(0.024)**	(0.023)*	(0.021)*	(0.023)**	(0.023)**	(0.020)**
Age 45 to 54	0.070	0.054	0.033	0.093	0.079	0.072
	(0.024)**	(0.024)*	(0.022)	(0.024)**	(0.023)**	(0.021)**
Age 55 to 65	0.070	0.040	0.036	0.047	0.024	0.076
	(0.027)**	(0.026)	(0.025)	(0.026)	(0.026)	(0.024)**
High School Graduate	0.054	0.030	0.060	0.105	0.093	0.063
	(0.031)	(0.031)	(0.029)*	(0.030)**	(0.029)**	(0.029)*
Some College	0.024	-0.005	0.025	0.112	0.066	0.085
	(0.032)	(0.032)	(0.030)	(0.031)**	(0.030)*	(0.030)**
College Graduate	0.048	0.008	0.054	0.136	0.053	0.112
	(0.036)	(0.033)	(0.033)	(0.033)**	(0.031)	(0.033)**
Post College	0.073	0.060	0.072	0.059	-0.066	0.055
	(0.042)	(0.037)	(0.038)	(0.044)	(0.035)	(0.040)
Male	0.064	0.053	0.041	-0.064	0.006	-0.068
	(0.022)**	(0.018)**	(0.019)*	(0.022)**	(0.018)	(0.019)**
Foreign Born	0.042	0.014	0.030	-0.137	-0.146	-0.109
	(0.027)	(0.027)	(0.026)	(0.028)**	(0.024)**	(0.021)**
Non-white	-0.170	-0.149	-0.172	-0.085	-0.045	-0.072
	(0.033)**	(0.033)**	(0.028)**	(0.035)*	(0.032)	(0.028)**
Year=1999	-0.025	-0.021	-0.003	-0.022	-0.035	-0.025
	(0.017)	(0.016)	(0.015)	(0.017)	(0.016)*	(0.015)
CPS Household Respondent	0.015	0.017	0.011	0.084	0.076	0.094
	(0.018)	(0.017)	(0.016)	(0.018)**	(0.017)**	(0.015)**
Self Emp Earnings $1000 to $9,999	0.116	0.135	0.217	0.057	0.098	0.149

	(0.034)**	(0.033)**	(0.024)**	(0.045)	(0.040)*	(0.027)**
Self Emp Earnings $10,000 to $49,000	0.248	0.251	0.364	0.068	0.148	0.280
	(0.031)**	(0.031)**	(0.022)**	(0.045)	(0.040)**	(0.027)**
Self Emp Earnings $50,000 or above	0.278	0.279	0.360	0.116	0.201	0.328
	(0.029)**	(0.029)**	(0.022)**	(0.046)*	(0.039)**	(0.030)**
Predicted Probability at Mean X	0.57	0.57	.053	0.62	0.60	.051

Appendix Exhibit 1
Self-Employment in the March CPS

Definition #1: We define an individual as self-employed in the March CPS based upon the following responses to the following questions:

>Q29a< Did you work at a job or business at any time during YYYY?
Yes

>Q46< What was your longest job during YYYY?

>Q47E1< [*At your longest job during YYYY*] Were you employed by government, by a PRIVATE company, a non-profit organization, or were you self employed or working in a family business?
Self employed or working in a family business

>Q47E1b< Was this business incorporated?
No

>Q47E1c< Are you the owner of the business?
Yes

>S-Q48a< How much did you earn from this employer before taxes and other deductions during YYYY?

Definition #2: We define an individual as self-employed in the March CPS based upon the following responses to the following questions:

If the longest job held during YYYY is not self employed or working in a family business:

>Q49a< Did you earn money from any other work (you/he/she) did during YYYY?
Yes

>S-Q49B1< How much did you earn from all other employers before taxes and other deductions during YYYY?

>Q49B2< How much did you earn from your own business after expenses?
Amount > 0

Question numbers are from the March 1995 Annual Demographic Survey March Supplement, accessed February 2006, from http://www.bls.census.gov/cps/ads/1995/sqestair.htm.

Appendix Exhibit 2
Self-Employment in the SSA Detailed Earnings Record

Self-Employment in the SSA Detailed Earnings Record

We define an individual as self-employed in the SSA Detailed Earnings Record if line 4 of IRS Form 1040 Schedule SE (Self-Employment Tax) is greater than $400.

The instructions for Schedule SE state "You must pay Self-Employment tax if you had net earning of $400 or more as a self-employed person. If you are in business for yourself or you are a farmer, you are self-employed. You must also pay SE tax on your share of certain partnership income and your guaranteed payments."

Below is a simplified version of Schedule SE (with slight wording changes):

1) Net profit or loss from Farms (Schedule F, line 36) and Farm Partnerships (Schedule K-1, Form 1065, box 14, code A) …….. _____

2) Net profit or loss from Business (Schedule C, line 31) and Business Partnerships (Schedule K-1, Form 1065, code A) ……...… _____

3) Combine lines 1 and 2 ……. _____

4) Net earnings from self-employment. Multiply line 3 by 92.35%. If less than $400, do not file this schedule; you do not owe self-employment tax …… _____

IRS Form 1040 Schedule SE is found at http://www.irs.gov/pub/irs-pdf/f1040sse.pdf (accessed February 2006).

Appendix Exhibit 3
Self-Employment in the Business Register

We define an individual as self-employed in the Business Register File if they are a sole proprietor (filing Schedule C).

The instructions for Schedule C state "Use Schedule C (Form 1040) to report income or loss from a business you operated or a profession you practiced as a sole proprietor. Also, use Schedule C to report wages and expenses you had as a statutory employee. An activity qualifies as a business if your primary purpose for engaging in the activity is for income or profit and you are involved in the activity with continuity and regularity. For example, a sporadic activity or a hobby does not qualify as a business."

Below is a simplified version of Schedule C (with slight wording changes):

Name of Proprietor …….…….. _____
SSN …… _____

Business Name (if no separate business name, leave blank) ……. _____
EIN (if any) …… _____

1) Gross receipts or sales …… _____

4) Cost of goods sold …….. _____

5) Gross Profits (subtract line 4 from line 3) …….. _____

7) Gross Income (add lines 5 and 6) …… _____

28) Total Expenses (add lines 8 through 27) ……..……. _____

31) Net profit or loss …… _____
If a profit, enter on Schedule SE line 2

IRS Form 1040 Schedule C is found at http://www.irs.gov/pub/irs-pdf/f1040sc.pdf (accessed February 2006).