UKCPR
University of Kentucky Center for Poverty Research

Discussion Paper Series
DP 2014-05

ISSN: 1936-9379

The Role of CPS Nonresponse on the Level and Trend in Poverty

Charles Hokayem
U.S. Census Bureau

Christopher Bollinger
Department of Economics
University of Kentucky

James P. Ziliak
Department of Economics
Center for Poverty Research
University of Kentucky

Preferred citation
Hokayem, C., Bollinger, C. & Ziliak,J., The Role of CPS Nonresponse on the Level and Trend in Poverty. *University of Kentucky Center for Poverty Research Discussion Paper Series, DP2014-05*.

Author correspondence
Charles Hokayem, 4600 Silver Hill Road, Washington, DC 20233; Email charles.hokayem@census.gov; Phone: (301)-763-5330

University of Kentucky Center for Poverty Research, 302D Mathews Building, Lexington, KY, 40506-0047
Phone: 859-257-7641; Fax: 859-257-6959; E-mail: ukcpr@uky.edu

www.ukcpr.org

EO/AA

The Role of CPS Nonresponse on the Level and Trend in Poverty*

Charles Hokayem, U.S. Census Bureau

Christopher Bollinger, Department of Economics, University of Kentucky

James P. Ziliak, Department of Economics and Center for Poverty Research, University of Kentucky

April 2014

Abstract: The Current Population Survey Annual Social and Economic Supplement (ASEC) serves as the data source for official income, poverty, and inequality statistics in the United States. There is a concern that the rise in nonresponse to earnings questions could deteriorate data quality and distort estimates of these important metrics. We use a dataset of internal ASEC records matched to Social Security Detailed Earnings Records (DER) to study the impact of earnings nonresponse on estimates of poverty from 1997-2008. Our analysis does not treat the administrative data as the "truth"; instead, we rely on information from both administrative and survey data. We compare a "full response" poverty rate that assumes all ASEC respondents provided earnings data to the official poverty rate to gauge the nonresponse bias. On average, we find the nonresponse bias is about 1.0 percentage point.

Key Words: ASEC, poverty measurement, hot deck imputation, nonresponse bias, inverse probability weighting
JEL Codes: I32 (Measurement and Analysis of Poverty)

---

*Contact: Charles Hokayem, U.S.Census Bureau, SEHSD, HQ-7H168, 4600 Silver Hill Rd, Washington, DC20233-8500. E-mail: charles.hokayem@census.gov Phone: 301-763-5330. James P. Ziliak, University of Kentucky, Lexington, KY 40506-0034; Email: jziliak@uky.edu; Phone: 859-257–2776. Christopher Bollinger, University of Kentucky, Lexington, KY 40506-0034; Email: crboll@uky.edu; Phone: 859-257-9524. We would like to thank Martha Stinson, Graton Gathright, and Gary Benedetto for their help in understanding the Detailed Earnings Record files. We thank Ed Welniak, Trudi Renwick, Chuck Nelson, Adam Bee, and session participants at the 2012 Joint Statistical Meetings, the 2013 Society of Labor Economists Meetings, 2013 Institute for Research on Poverty Summer Workshop, and University of Kentucky for helpful comments. We also thank the Summer at Census program for support. All comparative statements have undergone statistical testing, and unless otherwise noted, all comparisons are statistically significant at the 10 percent significance level.

The accurate measurement of income is vital to assessing economic growth, characterizing income poverty and inequality, gauging the effectiveness of the federal safety net, among other important applications. The Current Population Survey Annual Social and Economic Supplement (ASEC) serves as the official source of income and poverty statistics for the United States. ASEC respondents may be reluctant to answer income questions, or indeed any questions, out of concern for response confidentiality, or they may just have insufficient knowledge of the answers (Groves 2001). As seen in Figure 1, the nonresponse rate for ASEC earnings among workers (both item nonresponse and supplement nonresponse) has risen dramatically since the early 1990's. The earnings imputation rate has reached 20 percent (the line with squares), and nonresponse of the entire ASEC supplement adds an additional 10 percentage points to make total nonresponse about 30 percent in a typical year over the past decade (the line with diamonds).[1] Rates of item nonresponse for other earnings (e.g., self-employment) trended upward in the 1990's, but they only contribute 1-2 percentage points per year, implying most is due to wage and salary workers.[2] Because earnings accounts for over 80 percent of total income in national income accounts, failure to accurately measure it may significantly bias estimates of the income distribution.

This paper assesses whether and to what extent there is bias in official poverty rates caused by earnings nonresponse.[3] The poverty rate, which has been measured consistently since the late 1960s, is not only the key statistical barometer of the well being of low-income families in the U.S., but also is used in establishing the size of intergovernmental transfers for scores of

[1] The flag for whole supplement nonresponse is not well known, and as shown in Figure 1 ignoring supplement nonresponse omits 1/3 of all imputations.
[2] For the years 1987-2011, the average share of earnings nonresponse due to wage and salary earnings is about 95 percent while the average share of earnings nonresponse due to self-employment earnings is about 5 percent.
[3] Bias could also arise from nonresponse in other income sources used in constructing the poverty rate such as government transfer programs or private nonlabor income (e.g. retirement, rent/interest/dividends). Rates of nonresponse in the ASEC among these other income sources generally range from 0.5 to 4 percent depending on source, and thus are much less common than earnings imputation.

programs. For example, in 2006 the poverty rate was utilized by 39 federal programs to allocate billions of dollars of assistance to states, municipalities, and individuals (Gabe 2007). Thus, knowledge of potential bias from earnings nonresponse is important as it could have substantive budgetary implications.

The current approach of the U.S. Census Bureau is to retain earnings nonrespondents in the sample and to assign them earnings via a matched "donor" with similar demographic characteristics using a sequential "hot deck" procedure (Little and Rubin 2002). The advantage of this approach is that with weights the sample retains population representativeness, there may be efficiency gains from retaining the whole sample, and it is less subject to specification error found in model-fitting approaches (Andridge and Little 2010). However, the hot deck procedure may bias estimates of population statistics if the missing at random assumption does not hold (Bollinger and Hirsch 2013). Hirsch and Shumacher (2004) and Bollinger and Hirsch (2006) study the hot deck procedure in both the ASEC and the CPS Outgoing Rotation Group, and show the hot deck procedure causes earnings regression parameters to be biased. Given the bias in regression parameters there is a possibility the hot deck procedure could bias estimates of statistics derived from income such as poverty rates.

We propose a new approach to address the effect of earnings nonresponse on the level and trend in poverty. Similar to the hot deck approach we seek the missing counterfactual owing to nonresponse: what would the poverty rate be if all nonrespondents reported their earnings? We consider a full response poverty rate from the ASEC to be the ideal estimator, abstracting from questions about definitions of income and measurement error.[4] The ASEC has a long history, much of it characterized by high response rates. Our approach allows researchers to

[4] See, for example, Citro and Michael (1995), Ziliak (2006), Meyer and Sullivan (2012), and Short (2013) for discussion on income definitions and measurement error as pertains to poverty measurement.

track this long series and adjust for nonresponse as it worsens, rather than attempting to establish a new series.

To estimate what we call the "full-response poverty rate," we assemble a proprietary dataset of internal ASEC records matched to Social Security Detailed Earnings Records (DER) that covers survey years 1998-2009 and allows for the systematic study of long-term trends in income imputation and poverty rates.[5] The DER file contains earnings from all jobs reported on a worker's W-2 forms, including wages and salaries and income from self-employment subject to Federal Insurance Contributions Act (FICA) and/or Self-Employment Contributions Act (SECA) taxation. The DER data are central to our analysis; however, our procedure does not treat the administrative data as correct or the "truth," and in fact takes the rather surprising stance that the ASEC is the "ideal." While some research on wages has treated administrative records like the DER as correct (see Bound and Krueger (1991) and Bollinger (1998)), these analyses typically attempted to remove individuals whose characteristics (industry and occupation) were likely to indicate substantial under the table earnings. This approach will not work in estimation of poverty rates, since all individuals and families must be included. Moreover, recent research (Roemer 2002; Abowd and Stinson 2013) has suggested that this is not necessarily an appropriate approach, and in fact, the alleged "over-reporting" of CPS earnings among low-income persons may reflect actual earnings not in the DER such as unreported and/or uncovered earnings (both legal and illegal).

Two major issues arise in establishing the full-response poverty series: the DER data differ from the ASEC data, and not all nonrespondents are matched to the DER records. We can address these issues most simply by comparing ASEC poverty rates to DER poverty rates for

---

[5] The ASEC was matched to the DER in 1991 and 1994, and annually since 1996. With the CPS redesign in 1995 to all CATI/CAPI interviewing we begin with the 1998 survey year in order avoid changes in survey design affecting our analysis.

those who both report earnings to the ASEC and are matched to the DER. This provides a simple correction to the DER poverty rates for those who fail to report earnings in the ASEC. Similarly, we can compare ASEC poverty rates for those who are matched to the DER and those who are not. Doing so provides a correction to account for nonrespondents who are not matched. Together, the corrections provide an estimate of the "true" poverty rate that would emerge in the absence of earnings nonresponse. We also construct these corrections by demographic groups, thus allowing the correction to differ across groups with known differences in response rates.

We compare our approach to several alternatives, drawing out important differences in the various assumptions and the attendant implications for the level and trend in poverty. For example, Nicholas and Wiseman (2009, 2010) and Turek, et al. (2012), each examine the effect of nonresponse on poverty in select years using what we call a "plug-in" approach. That is, they replace ASEC earnings with DER earnings, or the maximum of ASEC and DER earnings, to construct an alternative poverty series, finding little effect of earnings nonresponse on poverty.[6] Our full-response poverty rate and simple alternatives such as the plug-in method make different assumptions about the underlying data generating mechanism, specifically whether data are missing completely at random, missing at random, or whether nonresponse is related to earnings (missing nonrandomly).

However, all of these approaches require access to the DER. Consequently, we also examine methods available to researchers without access to the DER in lieu of the hot deck. These approaches include Manski (1989) bounds, simply removing all nonrespondents, and

[6] In addition to our new method of accounting for nonresponse, our study differs from Nicholas and Wiseman (2009, 2010), and Turek, et al (2012), in several ways. First, we examine a longer time series. Second, unlike Turek, et al. who focus only on persons with positive earnings, we examine the entire poverty universe. Third, consistent with the Census construction of poverty as a family concept, we derive measures of nonresponse at the family level. Fourth, we distinguish the contribution of matched versus unmatched, respondent versus nonrespondent, and working versus nonworking families to the poverty rate.

inverse probability weights. As with the approaches using the DER, each non-DER alternative makes different sets of assumptions concerning the data generating process. By comparing these approaches to the results using the DER, we highlight methods which provide similar results and provide researchers with an approach which mimics the results obtained using the DER data.

Our results suggest that assumption of missing at random, even conditional on known characteristics, is not valid in modern data. Hence any correction which assumes missing completely at random or missing at random, such as the hot deck procedure, is likely to be biased. We show that the ASEC underestimates the number of persons in poverty by an average of about 1.0 percentage point.[7]

II. Poverty, Nonresponse, and Linked Administrative Data

The official poverty rate is based both on the actual earnings of those persons who respond to the ASEC earnings questions along with the imputed earnings of those persons who do not respond to the ASEC earnings questions. The poverty rate can be written as a weighted average of these two groups:

(1) $$P^C = P_R^{ASEC} * \Pr\{R\} + P_{NR}^{ASEC} * \Pr\{NR\},$$

where P^C is the official Census poverty rate, P_R^{ASEC} is the poverty rate among respondents (R) to ASEC earnings, $\Pr\{R\}$ is the probability of earnings response, P_{NR}^{ASEC} is the poverty rate among nonrespondents (NR) to ASEC earnings, and $\Pr\{NR\}$ is the probability of nonresponse. The ASEC data provide consistent estimates of three of the terms on the right hand side: P_R^{ASEC}, $\Pr\{R\}$, and $\Pr\{NR\}$. The term P_{NR}^{ASEC} is not identified in the ASEC data, and thus the current Census practice is to implement the hot deck procedure described in detail below to replace the missing earnings in order to derive an estimate of poverty. Our goal is to assess how earnings

[7] The Results Section (Section V) describes the test for this main finding.

nonresponse affects the official poverty rate in the ASEC. In the terminology of the program evaluation literature, we are missing the counterfactual—what would the poverty rate be if nonrespondents had responded to the earnings questions?

A point of departure for our analysis is to consider the extreme bounds cases motivated by the work of Manski (1989). Because the poverty rate falls between 0 and 1 we can place bounds on the official series by making the polar assumptions that the poverty rate among nonrespondents is 0 or the poverty rate among nonrespondents is 1. A nonrespondent poverty rate of 0 gives the lower bound (best case poverty scenario), or

(2) $$P^{LB} = P_R^{ASEC} * \Pr\{R\} + 0 * \Pr\{NR\} = P_R^{ASEC} * \Pr(R),$$

while a nonrespondent poverty rate of 1 gives the upper bound (worst case poverty scenario), or

(3) $$P^{UB} = P_R^{ASEC} * \Pr\{R\} + 1 * \Pr\{NR\} = P_R^{ASEC} * \Pr\{R\} + \Pr\{NR\}.$$

In Figure 2 we plot the upper and lower bounds against the official poverty rate. As can be seen, the lower bound is closer to the official rate than the upper bound. The lower bound differs from the official rate by about 3 percentage points. However, the upper bound is over three times the official rate. In both cases, they differ statistically from the official rate at the 5 percent level.

[Figure 2 here]

The Manski bounds are extreme because they assume we know nothing about the poverty status of nonrespondents. In fact, we know a lot about nonrespondents as we have linked administrative data on their earnings from the DER. With these data we seek to identify the missing counterfactual and establish a benchmark "full response" poverty rate, P^{Full}, that is more informative than the Manski upper bound. In constructing P^{Full} we must address the fact that earnings in the DER differ from those in the ASEC and that not all respondents in the ASEC have a linked record in the DER. Specifically, ASEC earnings reports can differ from DER reports both because not all jobs are covered by Social Security and thus not required to be recorded in the DER, and that "under the table" earnings could be reported to the ASEC that are not reported to the IRS. There is evidence that "under the table" earnings can occur in the low-income population (Edin and Lein 1997; Venkatesh 2006), and these earnings do not show up on tax data. Likewise, ASEC sample members will not be matched to the DER if (a) they did not give consent to be linked to the DER, (b) they had no earnings covered by the DER (legal or illegal), or (c) they did not work for pay. The former (difference between ASEC and DER reports) implies that we need to make an adjustment for measurement differences across the two series, while the latter (not matched to the DER) implies we need to make an adjustment for sample composition differences across the series.

Formally, we expand our decomposition of the poverty rate in equation (1) from two groups to four groups defined by respondent/nonrespondent status and DER match/nonmatch status as

	Match DER	Nonmatch DER
Respondent	$P_{R,M}^{ASEC} * \Pr\{R\&M\}$	$P_{R,NM}^{ASEC} * \Pr\{R\&NM\}$
Nonrespondent	$P_{NR,M}^{ASEC} * \Pr\{NR\&M\}$	$P_{NR,NM}^{ASEC} * \Pr\{NR\&NM\}$

where the subscript M refers to an ASEC sample member matched to the DER and NM is not matched to the DER. For example, in the top left cell, $P_{R,M}^{ASEC}$ is the poverty rate of respondents matched to the DER using ASEC earnings, and Pr{R&M} is the probability of responding to the ASEC *and* matched to the DER. This permits us to rewrite equation (1) as

(4) $P^C = P_{R,M}^{ASEC} * \Pr\{R\&M\} + P_{R,NM}^{ASEC} * \Pr\{R\&NM\} + P_{NR,M}^{ASEC} * \Pr\{NR\&M\} + P_{NR,NM}^{ASEC} * \Pr\{NR\&NM\}$.

We observe $P_{R,M}^{ASEC}$ and $P_{R,NM}^{ASEC}$ in the ASEC regardless of match status, and hereafter we collapse the first two terms in equation (4) as $P_R^{ASEC} * \Pr\{R\}$, which is simply the first term in equation

(1). However, we do not observe $P_{NR,M}^{ASEC}$ or $P_{NR,NM}^{ASEC}$, and thus use the DER earnings to provide an alternative measure of the earnings for these two unobservable poverty rates.

In equation (4), we replace $P_{NR,M}^{ASEC}$ with $P_{NR,M}^{DER}$, which is the poverty rate of matched nonrespondents using the DER as the measure of earnings. To account for measurement differences between the DER and the ASEC, we add a correction for measurement error among matched respondents: $(P_{R,M}^{ASEC} - P_{R,M}^{DER})$. Putting this together gives an estimator for the term $P_{NR,M}^{ASEC}$

(5) $\hat{P}_{NR,M}^{ASEC} = P_{NR,M}^{DER} + (P_{R,M}^{ASEC} - P_{R,M}^{DER})$.

We would like to make a similar substitution for $P_{NR,NM}^{ASEC}$ with $P_{NR,NM}^{DER}$ in the final term of equation (4), but we cannot measure the poverty rate of nonmatched nonrespondents using DER earnings, i.e $P_{NR,NM}^{DER}$ will never be observed. If we assume that nonmatched nonrespondents are similar to matched nonrespondents, we could use the estimator in equation (5) $P_{NR,M}^{DER} + (P_{R,M}^{ASEC} - P_{R,M}^{DER})$. However, the population who are not matched to the DER differs from those who are matched to the DER in both demographic characteristics and in earnings levels. To correct for these differences we compare the ASEC earnings of nonmatched respondents to matched respondents $(P_{R,NM}^{ASEC} - P_{R,M}^{ASEC})$. Substituting these expressions into the term for $P_{NR,NM}^{ASEC}$ gives our estimator

(6) $\hat{P}_{NR,NM}^{ASEC} = P_{NR,M}^{DER} + \left(P_{R,M}^{ASEC} - P_{R,M}^{DER}\right) + (P_{R,NM}^{ASEC} - P_{R,M}^{ASEC})$.

Our approach here allows for nonresponse to be related not only to demographic characteristics (as does the hot deck procedure), but also related to unobservable characteristics and the income level or poverty rate itself. The relationship to demographic characteristics is strengthened when we construct the poverty rate by weighted demographic subgroups below. Our approach also allows matching or failure to match to be related to demographic characteristics as well as unobservable characteristics. As such it also allows the DER and ASEC measures to differ and corrects for those differences. However, we assume that there is no interaction between these three mechanisms. That is, we are assuming that conditional on poverty status, measurement differences, nonresponse, and nonmatch are independent. In equation (5) the term $(P_{R,M}^{ASEC} - P_{R,M}^{DER})$ implies we are assuming that measurement differences between the DER and the ASEC do not differ between respondents and nonrespondents. That is, if nonrespondents were to respond, the differences between their DER record and their ASEC response would be similar to the differences between current respondents DER and ASEC. In equation (6) the term $(P_{R,NM}^{ASEC} - P_{R,M}^{ASEC})$ implies we are assuming that the differences in poverty rates between the matched and nonmatched populations are the same in both the DER and the ASEC. The first set of assumptions, which allow equation (5) to provide an estimate of the term $P_{NR,M}^{ASEC}$ are weaker than the missing at random assumption used in the hot deck procedure. Indeed, if the missing at random assumption holds, the results in equation (5) should be equivalent (up to sampling error) with using the hot deck procedure. The second set of assumptions regarding the match are not required for the hot deck procedure since the hot deck does not involve matching to the DER. However, if missing at random holds, and nonmatched at random were also to hold, then again our procedure should be similar to the hot deck. Our procedure allows both of these assumptions to fail, but requires nonmatch, nonresponse, and measurement error processes to be independent of each other, conditional on poverty status. Substituting (5) and (6) into (4) gives our benchmark expression for the full response poverty rate, P^{Full}, as

(7) $$P^{Full} = P_R^{ASEC} * \Pr\{R\} + \left(P_{NR,M}^{DER} + \left(P_{R,M}^{ASEC} - P_{R,M}^{DER}\right)\right) * \Pr\{NR\&M\} + \left(P_{NR,M}^{DER} + \left(P_{R,M}^{ASEC} - P_{R,M}^{DER}\right) + (P_{R,NM}^{ASEC} - P_{R,M}^{ASEC})\right) * \Pr\{NR\&NM\}.$$

It is worth emphasizing that the expression in (7) consists solely of *observed* data—both survey and administrative—and thus serves as our estimate of the "true" poverty rate in the U.S.

Against the benchmark in (7) we compare simpler alternatives utilizing the DER, what we call "plug-in" estimates of poverty. For example, we can replace ASEC earnings with DER earnings for all persons with a DER match regardless of imputation status and use ASEC (hot deck) earnings for persons without a DER match:

(8) $$P_{R,NR}^{Plug-in} = \boldsymbol{P_{R,M}^{DER}} * \Pr\{R\&M\} + P_{R,NM}^{ASEC} * \Pr\{R\&NM\} + \boldsymbol{P_{NR,M}^{DER}} * \Pr\{NR\&M\} + P_{NR,NM}^{ASEC} * \Pr\{NR\&NM\},$$

where the DER-based earnings poverty rates are highlighted in bold. This approach implicitly assumes that survey reports in the ASEC are mismeasured and the DER records provide a superior measure of earnings. This may not be true, however, both because some earnings reported in CPS are not taxable, and some earnings may be reported to the Census but not to the IRS, especially self-employment earnings and "under the table" earnings (Bound and Krueger 1991; Bollinger 1998; Roemer 2002).

In a related approach we replace ASEC earnings with DER earnings only for those nonrespondents with a DER match, and use reported ASEC earnings for respondents and (hot deck) ASEC earnings for persons without a DER match:

(9) $$P_{NR}^{Plug-in} = P_{R,M}^{ASEC} * \Pr\{R\&M\} + P_{R,NM}^{ASEC} * \Pr\{R\&NM\} + \boldsymbol{P_{NR,M}^{DER}} * \Pr\{NR\&M\} + P_{NR,NM}^{ASEC} * \Pr\{NR\&NM\}.$$

The logic here is that DER earnings for the actual worker dominate imputed earnings from an unrelated person, especially if the missing at random assumption is violated (either because the imputation algorithm uses too sparse a set of demographics, or there is selection on unobservables).

The approaches in equations (8) and (9) are closely related to Turek, et al. (2012) and Nicholas and Wiseman (2009, 2010). Turek, et al. merge earnings information from the DER to the 2006 ASEC (calendar year 2005) to examine the effect of substituting DER earnings for reported ASEC earnings on income estimates and number of persons in poverty. They focus only on workers, and those with a DER match, and thus do not examine the entire poverty universe as we do here. Nicholas and Wiseman (2009, 2010) merge both DER and Supplemental Security Income (SSI) administrative data to the ASEC to study poverty among elderly persons in the U.S. (and also present a subset of estimates for the whole population). Their approach is a variant of equation (8) where they use the maximum of the ASEC and DER for the measure of earnings and employ a reweighting adjustment for ASEC observations unmatched to the administrative data. This method makes the strong assumption that measurement error in the ASEC is always negative (not simply an underreport of true earnings on average, but never an overreport). This is particularly strong when considering the nonrespondents. In these cases the DER is used only when it exceeds the hot deck imputation. Since the hot deck is a random match, we expect it to contain differences that are both positive and negative. It also makes the somewhat weaker assumption that earnings in the DER are always an understatement of true earnings. Since the procedure also uses the hot deck earnings for individuals who are not matched to the DER, this assumes that nonmatched, nonrespondents are missing and unmatched at random. This approach, by construction, will necessarily result in a lower poverty rate than that achieved by the hot deck procedure. There is little justification for a procedure designed this way. Finally, this series will differ from the historical ASEC series, since it is considering an income definition that differs from both the survey and DER measures. While in some ways this may be an improved measure of poverty, our approach seeks to provide a series that would be comparable to the measures in historical ASEC poverty rates (when nonresponse was negligible) and to other data series with smaller nonresponse rates. However, for completeness we will present a third plug-in poverty estimate, $P_{Max}^{Plug-in}$, based on the maximum ASEC/DER report.

III. Measuring Poverty in the Presence of Nonresponse without Administrative Data

The linked ASEC-DER data are not available in the public domain since they consist of proprietary tax information, and thus the research community at large cannot construct the full-response poverty rate in equation (7). Indeed, the linked data are only available with a lag, and thus if utilized by the Census Bureau could result in costly delays in releasing official poverty statistics. This then begs the question: what do you do in the absence of administrative data to handle nonresponse? If we assume that earnings are missing completely at random (MCAR), then the best strategy is to simply drop nonrespondents altogether and report P_R^{ASEC} as the official poverty rate. Like Manski bounds, this too may be extreme because of evidence that nonresponse is correlated with several observable characteristics such as education, gender, and race (Bollinger and Hirsch 2006, 2013). The primary characteristic in which they differ is that all nonrespondents are earners. Since, on average, earners are less likely to be in poverty than households with no one in the labor force, the sample is then no longer representative, and indeed is biased toward a higher poverty population than the population as a whole. We have substantial demographic information on nonrespondents in the ASEC. Indeed, even in the case of whole supplement nonrespondents, we have some information on the basic demographic profile of the household. Thus a more robust approach is to assume the data are missing at random, and below we assess several alternatives to the full response poverty rate in (7).

A. The ASEC Hot Deck Imputation Procedure

The Census Bureau makes the missing at random (MAR) assumption to utilize a set of demographic variables to match nonrespondents to respondents, and thus to replace the missing earnings of the nonrespondent with the earnings of the matched respondent to implicitly construct an estimate of P_{NR}^{ASEC}. Specifically, Census has used a hot deck procedure for imputing missing income since 1962, and the current system has been in place with few changes since 1989 (Welniak 1990). The cell hot deck procedure assigns individuals with missing earnings values that come from individuals with similar characteristics. The ASEC uses a variation of the cell hot deck procedure known as a sequential match procedure. First, individuals with missing data are divided into one of 12 allocation groups defined by the pattern of nonresponse. Examples include a group that is only missing earnings from longest job or a group that is missing both longest job information and earnings from longest job. Second, an observation in each allocation group is matched to another observation with complete data (called the donor) based on a large set of socioeconomic variables, the match variables.[8] If no match is found based on the full set of match variables, then a match variable is dropped and variable definitions are collapsed to be less restrictive. The process of sequentially dropping variables and collapsing variable definitions is repeated until a match is found. When a match is successful, the missing earnings is substituted with the reported earnings from the first available donor.

The ASEC also uses a hot deck procedure for whole supplement, or unit, nonresponse. In this context, whole imputation refers to an individual who responds to the monthly basic earner study but does not respond to the ASEC supplement and requires the entire supplement to be imputed. Instead of 12 allocation groups, the whole imputation procedure uses 8 allocation

---

[8] The set of match variables includes gender, race, age, relationship to householder, years of school completed, marital status, presence of children, labor force status of spouse, weeks worked, hours worked, occupation, class of worker, other earnings receipt, type of residence, region, transfer payments receipt, and person status.

groups. Moreover, the set of match variables is smaller than the set used for item nonresponse, consisting solely of variables from the basic monthly CPS. To be considered a donor for whole imputations, an ASEC respondent has to meet the minimum requirement that at least one person in the household has answered one of the following questions: worked at a job or business in the last year; received federal or state unemployment compensation in the last year; received supplemental unemployment benefit in the last year; received union unemployment or strike benefit in the last year; or lived in the same house one year ago. This requirement implies that whole supplement donors do not have to answer all the ASEC questions and can have item imputations. Similar to the sequential hot deck procedure for item nonresponse, the match process sequentially drops variables and makes them less restrictive until a donor is found.

B. Inverse Probability Weighting

A potential pitfall of the Census hot deck procedure is the finite set of covariates that are used to find a matched donor. An alternative approach for nonresponse under the MAR assumption is inverse probability weighting (IPW) (Wooldridge 2007, 2010). IPW is likely to be more general than the hot deck procedure as it offers a solution to the "curse of dimensionality," which arises from the computational burden with hot deck from expanding the set of covariates used to match. As such, IPW is closely related to the propensity score method in the treatment effects literature (Rosenbaum and Rubin 1983). IPW, like the hot deck procedure, assumes missing at random. It further assumes that a set of demographic factors, $\boldsymbol{z_i}$, are observed that predict whether the individual responds or not to the earnings questions, and that the resulting prediction of response probability is everywhere nonzero.[9] With these assumptions we can obtain a consistent estimate of the population poverty rate (Wooldridge 2010, p. 822-823)

(10) $P = \sum_{i=1}^{n}(R_i P_i / \Pr\{\boldsymbol{z_i}\}) / \sum_{i=1}^{n}(R_i / \Pr\{\boldsymbol{z_i}\})$,

which is the poverty rate of respondents weighted by the inverse probability of response. To implement the IPW approach we need to fit a flexible model of the probability of response, $\Pr\{R\}$, which can include higher-order powers and interactions of the $\boldsymbol{z_i}$. The computational advantage of IPW over the hot deck then becomes clear because it is not necessary to find a donor that matches across a wide array of characteristics, but only a single index of the probability of response.

In practice, the Census Bureau weights the ASEC sample poverty rate by the inverse probability of sample inclusion adjusted for survey nonparticipation and other special factors in order to be representative of the U.S. population.[10] Thus, the current official poverty rate is an IPW estimator; that is, $P^C = \sum_{i=1}^{n} w_i P_i \,/\, \sum_{i=1}^{n} w_i$, where w_i is the (adjusted) inverse probability of sample inclusion for the individual and $\sum_{i=1}^{n} w_i$ is an estimate of the U.S. population. This means that if we wish to retain population representativeness utilizing the ASEC for respondents only, then we need to adjust the Census measure for earnings nonresponse as

(11) $P^{IPW} = \sum_{i=1}^{n}(R_i P_i w_i / \Pr\{\boldsymbol{z_i}\}) / \sum_{i=1}^{n}(R_i w_i / \Pr\{\boldsymbol{z_i}\})$,

which weights up the respondent sample so that the population estimate is retained. For example, if $w_1 = 10{,}000$ (i.e. the probability of person 1 being sampled is 0.001 so that they represent 10,000 people), and the probability of person 1 being a respondent is $\Pr\{\boldsymbol{z_1}\}$ =0.4, then

[9] Strictly, for persons out of the labor force such as retirees we can identify their poverty status even if they are an earnings nonrespondent. Thus in implementing the various poverty estimators we will differentiate between working and non-working individuals.

[10] See Current Population Survey, Design and Methodology, Technical Paper 66, http://www.census.gov/prod/2006pubs/tp-66.pdf (accessed October 30, 2013) for details.

the new weight for person 1, $w_1/\Pr\{z_1\}$, is 25,000. Below we compare both the official poverty rate and the IPW variant based on respondents only to our benchmark measure.

[Table 1 here]

Table 1 summarizes the differences in assumptions between the various approaches examined in this paper. The Manski bounds have the weakest set of assumptions, but necessarily yield the weakest conclusions. Manski bounds are useful in assessing the uncertainty inherent with the problem of missing data, and also in evaluating estimators with differing assumptions. In this research we are using the CPS measures of income as "correct" not because we strongly believe they are superior, but rather because they represent a common (over time and across surveys) measure. Hence, we argue that any approach which rejects this assumption does not meet the spirit of the endeavor (specifically the all DER measures and the Max measure). Our approach, P^{Full}, is one of the three approaches which only use two assumptions (one being that the CPS is correct). The official hot deck approach, P^{C}, and inverse probability weighting, P^{IPW}, both only assume missing at random and that the CPS is correct. If missing at random holds, our approach will also work provided our independence assumption holds as well. However, our results below, the work of Bollinger and Hirsch (2013), and recent work by Hokayem et al (2013) provide evidence that missing at random does not hold. Hence the IPW and hot deck are likely to be biased, and our approach uses the weakest assumptions besides the Manski bounds.

IV. Data

Our sample consists of the entire Census poverty universe; that is, all noninstitutionalized families and unrelated individuals ages 15 and older from the ASEC for survey years 1998-2009 (reporting income for 1997-2008). The ASEC is then matched to the Social Security

Administration's Detailed Earnings Record (DER) file.[11] The DER file is an extract of Social Security Administration's Master Earning File (MEF) and includes data on total earnings, including wages and salaries and income from self-employment subject to Federal Insurance Contributions Act (FICA) and/or Self-Employment Contributions Act (SECA) taxation. Since individuals do not make SECA contributions if they lose money in self-employment, only positive self-employment earnings are reported in the DER file (Nicholas and Wiseman 2009). Nonworkers and those who do not pay into Social Security are not in the DER.

Workers in the DER file are uniquely identified by a Protected Identification Key (PIK) assigned by the Census Bureau, which is a confidentiality-protected version of the Social Security Number (SSN). The Center for Administrative Records Research and Applications (CARRA) within Census matches the DER file to the ASEC. Since the CPS does not currently ask respondents for a SSN, CARRA uses its own record linkage software system, the Person Validation System, to assign a SSN.[12] This assignment relies on a probabilistic matching model based on name, address, date of birth, and gender (NORC 2011). The SSN is then converted to a PIK, and the SSN from the DER file received from SSA is also converted to a PIK. The ASEC and DER files are matched based on the PIK and do not contain the SSN.

The DER file contains earnings reported on a worker's W-2 form, and there is one record for each W-2 for those workers holding multiple jobs. Figure 3 provides a sample W-2 form with the circled boxes we use in the analysis. These earnings are not capped at the FICA contribution amounts and include earnings not covered by Old Age Survivor's Disability

[11] The DER were not linked to the 2001 SCHIP expansion sample in 2001, but were in each year thereafter. For more information on sampling and non-sampling error in the ASEC, see http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf (accessed October 30, 2013).

[12] The final year the CPS collected respondent Social Security Number is CPS survey year 2005 (calendar year 2004). Beginning with survey year 2006 (calendar year 2005), all respondents were assigned a Social Security Number using the Person Validation System, whereas in prior years a SSN was assigned only if the SSN was not valid.

Insurance (OASDI) but subject to Medicare tax. Unlike ASEC earnings which are top-coded by Census, DER earnings are not top-coded. The DER file also contains deferred wages such as contributions to 401(k), 403(b), 408(k), 457(b), 501(c), and HSA plans, but not pre-tax health insurance premiums and education benefits and thus is not a complete source of gross compensation (Abowd and Stinson 2013). Since a worker can appear multiple times per year in the DER file if they have several jobs, we collapse the DER file into one earnings observation per worker per year by aggregating total compensation (Box 1 of W-2), SSA covered self-employment earnings (SEI-FICA), and Medicare covered self-employment earnings (SEI-MEDICARE) across all employers. DER earnings are defined as the sum of total compensation plus the maximum of SSA covered self-employment income or Medicare covered self-employment:

DER Earnings = (Box 1 of W-2) + max(SEI-FICA,SEI-MEDICARE).

In this way DER earnings is most compatible with the CPS earnings (PEARNVAL), which cover earnings from all wage and salary jobs (WSAL-VAL), business self-employment (SEMP-VAL), and farm self-employment (FRSE-VAL). The CPS total personal income variable (PTOTVAL) used to determine poverty status consists of adding a person's total earnings to a person's total other income (POTHVAL), PTOTVAL=PEARNVAL+POTHVAL.

[Figure 3 here]

Like the match to the DER, imputations of earnings occur at the individual level as well. For our purposes we classify a worker as having imputed earnings if either wages and salary from longest job is imputed (I-ERNVAL), wages and salary from other jobs is imputed (I-WSVAL), self-employment earnings is imputed (I-SEVAL and I-FRMVAL), or the whole ASEC supplement is imputed (FL-665). However, since the official poverty rate in the U.S. is a family concept, Census sums individual income across all persons in the family to create family income that is compared to the official poverty threshold. Thus, to be consistent with the family definition of poverty, we aggregate individual income nonresponse and match status to create family level variables. That is, a family is considered imputed if *any* member in the family has imputed earnings, or has the entire supplement imputed. A family is considered matched to the DER data if *all* earners in the family are matched to a DER record. An implication is that it is possible for families to contain no workers, especially among retirees and the disabled, and thus by construction no match with the DER is possible for the family. Consequently we have to modify our full poverty rate in equation (7) to be conditional on earner status in the family, i.e

(12) $P^{Full} = (P^{Full}|earner \geq 1)\Pr\{earner \geq 1\} + (P^{Full}|earner = 0)\Pr\{earner = 0\}.$

These new definitions of family imputation and family match differ from previous research (e.g. Nicholas and Wiseman 2009, 2010; Turek, et al. 2012) that defined these concepts at the individual level, which is incongruent with the family-level construct of Census poverty. We note, however, that the level and trend in earnings nonresponse (item and whole supplement) is qualitatively little affected by aggregating up to the family level compared to the individual level in Figure 1, rising from 26.4 percent in 1997 to 29.4 percent in 2008.[13] Figure 4 depicts trends in the family level ASEC-DER match rate conditional on earners in the family (recall that by construction a family cannot be matched if there are no earners). In 1997 just over 60 percent of earner families in the ASEC were matched to the DER, and this rose to 74 percent starting in 2005 and held steady thereafter.[14] The shift up in matches most likely occurred because Census changed the consent process for linking to SSA data from "opt-in" to "opt-out," i.e. starting in 2005 sample members were automatically enrolled in the link process and had to request that

[13] The nonresponse rate in 1997 is statistically different from the nonresponse rate in 2008.
[14] The match rate in 1997 is statistically different from the match rate in 2005.

they be removed. Importantly there is a 20-25 percentage point difference in DER match rates depending on whether the family is a respondent or nonrespondent, highlighting the importance of distinguishing match/nonmatch by respondent status in our full-response poverty rate in equation (7).[15]

[Figure 4 here]

V. Results

Table 2 presents detailed summary statistics of the sample family head based on match and respondent status. Across most demographic characteristics the differences between matched and nonmatched families is much more pronounced among respondents than nonrespondents. For example, matched respondents are 14 years younger on average than nonmatched respondents, reflecting the fact that the latter group is much more likely to be retired or disabled, while there is only a 3 year age gap between matched and nonmatched nonrespondents. In both cases, the differences are statistically significant at the 5 percent level. Likewise matched respondents are statistically much less likely to be a high school dropout or to be living in poverty than nonmatched respondents. These gaps are relatively small among nonrespondents. As a consequence, ASEC earnings and family income are substantially and statistically higher for matched than nonmatched respondents. Interestingly, though, is that even though the difference in earnings and income among nonrepondents is comparatively small across match status, the level is higher than among respondents, suggesting that high income persons are less likely to respond to the ASEC.[16]

---

[15] The match rate for respondent families is statistically different from the match rate for nonrespondent families in all years. We note that the family level DER match rate of earners is about 10-12 percentage points lower than individual level match rates. This occurs because about 10 percent of families have more earners than DER matches, and thus we classify the whole family as nonmatched.

[16] ASEC and DER family earnings and income for matched respondents is statistically different from ASEC and DER family earnings and income for matched nonrespondents. ASEC family earnings and income for nonmatched respondents is statistically different from ASEC family earnings and income for nonmatched nonrespondents.

[Table 2 here]

Table 3 presents our benchmark full-response poverty estimates from equations (7) and (12), and compares it to the other poverty rates based on the administrative DER data: the two "plug-in" poverty rates using the DER, and the third that uses the maximum earnings between the ASEC and DER for matched nonrespondents. The full response poverty rate ranges from a low of 11.9 percent in 2000 to 14.2 percent in 2008, and these differences are statistically significant. The plug-in poverty rate using the DER for both respondents and nonrespondents from equation (8) is higher than the benchmark rate in all but one of the years where they are statistically different. Just the opposite occurs with the plug-in rate using DER for nonrespondents only in equation (9), where the series is statistically much lower in each year. This is even more pronounced when using the maximum of the ASEC and DER earnings, which was expected because this method makes the strong assumption that measurement error in the ASEC is always negative and thus will result in a poverty rate lower than the full response rate (and the official rate). Taken together, these results suggest that the missing at random assumption does not appear to hold: nonrespondents are more likely to be in poverty than their matched counterparts.

[Table 3 here]

In Tables 4a and 4b we explore in finer detail the components of the full response poverty rate that might shed light on why the official poverty rate is systematically lower. In Table 4a we present the components of equation (7) for families with at least one earner (recall equation (12) where the full response poverty rate is the weighted sum of the rates of earner and nonearner families). The numbers in bold in columns (1), (4), and (6) sum up to the number in column (7) subject to rounding error. Of particular note is column (3) where we compare the poverty rates of matched respondents using ASEC earnings versus DER earnings. The difference is negative, which means that ASEC earnings are higher and poverty rates lower than in the DER, suggesting that ASEC earnings captures income sources not reported to the DER either because they are not taxable or they are "under the table."[17] On the other hand, in column (5) we report the difference in poverty rates of nonmatched respondents and matched respondents in the ASEC. This difference is statistically significant and positive, suggesting that nonmatched respondent families are systematically poorer than matched families. This correction grows over time, especially after 2004 when the Census changed from the "opt-in" to the "opt-out" consent of being linked. In Table 4b we present the same calculations for nonearner families. Note that most of the terms are zero since by construction nonearner families are not matched to the DER. Also notable is the fact that the poverty rates of nonearner families are more than three times higher that earner families.[18] The full poverty rate reported in Table 3 is much closer to the earner rates because the probability of a family containing at least one worker averages over 85 percent in each year so that the earner sample receives nearly 6 times more weight in the full poverty calculation.

[Tables 4a-4b here]

In Table 5 we compare the benchmark full poverty rate to poverty rates based solely on publicly available ASEC data, including the poverty rate of ASEC respondents only under the missing completely at random assumption (P_R^{ASEC}), the official poverty rate under missing at random (P^C), and the IPW poverty rate also derived under missing at random (P^{IPW}). The poverty rate from respondents only is systematically too high by 0.4 percentage points on average, suggesting that the missing completely at random assumption is incorrect for missing

[17] The poverty rate of matched respondents using ASEC earnings is statistically different from the poverty rate of matched respondents using DER earnings in all years.

[18] The poverty rate of nonearner families is statistically different from the poverty rate of earner families in all years.

earnings reports.[19] On the other hand, the official poverty rate is statistically significantly lower in each year, averaging about 1.0 percentage point lower than the full-response benchmark, and this gap seems to have widened over time. This suggests that the official rate is undercounting poverty compared to a rate in which all sample members respond to the earnings questions.

[Table 5 here]

For the IPW poverty rate, we examined combinations of three different sets of demographic characteristics and three different models for the probability of nonresponse. The demographic characteristics include age (represented by a quartic function), race, gender, education (represented by degree and attendance indicator variables), marital status, class of worker (private, federal, state or local), foreign born citizenship status, and occupation. At the family level we also measure the size of metropolitan area, the region of the country, and the respondent's relationship to the head of the household. We measure the individual demographic characteristics for the head of the family, the spouse of the head (suppressing the spouse's marital status), and the respondent for the household. We estimate models using head's characteristics only, head and spouse, head and respondent, and all three sets always in conjunction with the household level characteristics. We used a linear probability model estimated using OLS, as well as probit and logit specifications estimated using maximum likelihood. We find virtually no qualitative differences between the final estimated IPW poverty rates across these specifications, and in Appendix Tables 1a and 1b we report the coefficients from the probit specification. We further experimented with dropping the respondent's relationship to the head of the household in models including only the head of the household characteristics. We find that the resulting estimated poverty rates do not differ either statistically or qualitatively across any of these specification choices. Since the respondent is not identified

[19] The poverty rate from respondents only is statistically different from the full poverty rate in all years but 2007.

in public use data in years prior to 1999, we settle on a probit model using only the characteristics of the head of the household and the household variables excluding the relationship of the respondent to the head. The models are estimated year by year.

With the estimated parameters we reweight the poverty rate of respondents in equation (11) and find that the IPW poverty rate is statistically and qualitatively systematically lower than the benchmark by about 0.5 percentage point. We note that if we take the simple average of the respondent-only poverty rate and the IPW poverty rate based on measured earnings in the ASEC (last two columns of Table 5) we derive a poverty series that is not statistically different from the full response poverty rate in the majority of years, and when they do differ they are sometimes higher and sometimes lower.[20] The advantage of the average of the respondent poverty rate and the IPW rate is that they are obtained from public release ASEC data and do not require the proprietary DER. However, understanding the differences between the respondent-only poverty rate and the IPW poverty rate obtained from the public release ASEC data is challenging.

[Figure 5 here]

To better understand why the respondent-only poverty rate is higher, Figure 5 presents the nonresponse rate by DER earnings percentile for persons ages 18-64, not enrolled in school full-time, and not in the Armed Forces. This sample includes workers and nonworkers. It shows nonresponse is relatively flat throughout most of the distribution with a spike at the top. Approximately 85 to 87 percent of the sample consists of “earner” families. Nonearner families have a higher poverty rate (32.3 to 37 percent) than families with at least one earner (8.8 to 9.4 percent) (last columns of Tables 4a and 4b). Nonearner families, by definition, have a zero nonresponse rate. A typical nonrespondent family has a much lower probability of being in

[20] The average of the respondent poverty rate and the IPW poverty rate is statistically different from the full response poverty rate only in 2001, 2003, and 2007.

poverty than a typical earner family. The respondent-only poverty rate that “drops all nonrespondents” disproportionately drops families that are less likely to be in poverty and is exacerbated by dropping families at the top end of the distribution. This causes the approach of “dropping all nonrespondents” to overstate the poverty rate because the sample is biased away from families with workers.

[Figure 6 here]

While the respondent-only approach overstates the poverty rate, the IPW approach understates the poverty rate. Figure 6 helps understand why the IPW approach has this effect. To obtain this figure we estimate a log DER wage equation by gender using the covariates in the inverse probability weighting models and in the hot deck procedure. Note this wage equation uses DER earnings rather than CPS earnings, and is for workers only. Figure 6 plots the nonresponse rate across the residual distribution from the wage equation by gender. It shows a pronounced U-shape, with high nonresponse in both tails of the residual distribution. Even after conditioning on the covariates typically used in the hot deck procedure and in the IPW approach, we still see a double selection found in the extreme tails of the residual distribution. This suggests a violation of the missing at random assumption necessary for both the hot deck procedure and the IPW approach. This also suggests there are unobservables causing nonresponse among individuals with low earnings than what observables would predict. The hot deck procedure and the weighting adjustments in the IPW approach do not account for these unobservables.

[Tables 6 and 7 here]

We conclude our empirical analysis by presenting the full response and official poverty rates for a variety of demographic groups to examine whether there are particular subsamples driving the results. Table 6 shows that on average families with children have full poverty rates 1.4 percentage points higher than the official rate, families headed by a female have full rates 1 percentage point higher, and families headed by a nonwhite or nonblack (other race) have full rates 1.5 percentage points higher on average.[21] It does appear that some groups have a larger effect on the full poverty series than others. To further examine if this heterogeneity is suppressed in our full poverty rate of equations (7) and (10), we compute a demographically weighted full poverty rate as $P_{Demog}^{Full} = \sum_{j=1}^{J} s_j \, P_j^{Full}$, where s_j is the population share of group j and P_j^{Full} is the subgroup j full response poverty rate. We use 18 groups formed by the partition of race (black, white, other race), headship (female, male), and education (less than high school, high school, and more than high school). This approach allows the correction to differ across groups with known differences in response rates. We report this approach in Table 7 where we see that there is no substantive or statistical difference in the full poverty rate and demographically weighted rate in most years.[22]

VI. Conclusion

This paper uses a unique dataset of administrative earnings data matched to internal ASEC to study the effects of earnings imputation on poverty measurement. Our analysis estimates the bias caused by earnings nonresponse. We compare a "full response" poverty rate that assumes all ASEC respondents provided earnings data to the official poverty rate to gauge this bias. On average, we find the nonresponse bias to be about 1.0 percentage point. This bias

---

[21] The difference between the full poverty rate and the official poverty rate for children is statistically significant in all years but 2000. The difference between the full poverty rate and the official poverty rate for families headed by a female is statistically different in all years but 1997 and 2000. The difference between the full poverty rate and the official poverty rate for families headed by a nonwhite or nonblack (other race) is statistically significant in all years but 1997 and 1998.

[22] The only years where there is a statistical difference between the full poverty rate and the demographically weighted poverty rate are 2002, 2003, and 2007.

seems more pronounced among more economically disadvantaged groups such as single female-headed families and those families headed by a nonwhite.

Our study is somewhat unique in that we take the stance that earnings reported in the ASEC are "ideal" compared to administrative reports in the DER. This stems from the fact that not all earnings are subject to Social Security taxation and thus not reported in the DER, and "under the table" earnings may show up in the ASEC but are not reported to tax authorities. This seems borne out in our sample in that poverty rates across the 12 years among matched respondents averages a statistically significant 1.7 percentage points lower using ASEC earnings than DER earnings. This suggests that simply replacing ASEC earnings with DER earnings is not the best solution to earnings nonresponse.

However, even though ASEC earnings may be preferred to DER earnings, simply dropping nonrespondents is not ideal either. Our estimates suggest that dropping nonrespondents results in a poverty rate systematically higher than our preferred full-response poverty rate. The bias caused by dropping nonrespondents likely stems from the loss of high earners who overall are more likely to be nonrespondents (see Bollinger and Hirsch, 2013). Moreover, Little and Rubin (2002) make a compelling case against such practice because of the potential loss of efficiency and representativeness. To address the latter concern, we constructed an inverse probability weighted poverty series and found that this series results in too low of a poverty rate relative to our benchmark (typically in the middle of the range between the official poverty rate and the full-response rate). On the other hand, a non-structural simple average of the poverty rates from dropping nonrespondents and the inverse probability weighted series is qualitatively and statistically no different than our full response series in most years. Thus, while we argue that our preferred measure of poverty is the full response measure, it is not possible for researchers and analysts to construct it without access to the DER, so the simple average may be a fruitful alternative since it relies solely on publicly available data in the ASEC. The weighted average is a compromise, but is more serendipitous than structural. As such, it should be corroborated in future samples and possibly in previous years. Most importantly, however, the accuracy of official poverty estimates in the U.S. would benefit greatly from reduced nonresponse of earnings and other income sources.

References

Abowd, John and Martha Stinson. Forthcoming. “Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data.” *Review of Economics and Statistics*, 95(5), pp 1451-1467.

Andridge, Rebecca, and Roderick Little. 2010. “A Review of Hot-Deck Imputation for Survey Nonresponse.” *International Statistical Review*, 78(1), pp 40-64.

Bollinger, Christopher. 1998. “Measurement Error in the CPS: A Nonparametric Look.” *Journal of Labor Economics*, 16(3), pp 576-594.

Bollinger, Christopher and Barry Hirsch. 2013. “Is Earnings Nonresponse Ignorable?” *Review of Economics and Statistics*, 95(2), pp. 407-416.

Bollinger, Christopher and Barry Hirsch. 2006. “Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching.” *Journal of Labor Economics*, 24(3), pp. 483-519.

Bound, John and Alan B. Krueger. 1991. “The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?” *Journal of Labor Economics* 9(1), pp. 1-24.

Citro, Constance, and Robert Michael. 1995. *Measuring Poverty: A New Approach*, Washington, D.C.: National Academy Press.

Edin, Kathryn, and Laura Lein. 1997. *Making Ends Meet: How Single Mothers Survivie Welfare and Low-Wage Work*. New York: Russell Sage Foundation.

Gabe, Thomas. 2007. “Federal Programs that Use the “Official” Poverty Definition for Determining Eligibility or for Allocating Funds,” Memo from the Congressional Research Service to the U.S. House Ways and Means Subcommittee on Income Security and Family Support (October 30).

Groves, Robert. 2001. *Survey Nonresponse*. New York: Wiley.

Hirsch, Barry and Edward Schumacher. 2004. “Match Bias in Wage Gap Estimates Due to Earnings Imputation.” *Journal of Labor Economics*, 22(3), pp. 689-722.

Hokayem, Charles, Christopher R. Bollinger, Barry Hirsch, and James P. Ziliak. “Trouble in the Tails? Earnings Nonresponse and Response Bias across the Distribution Using Matched Household and Administrative Data.” Mimeo, U.S. Census Bureau, October 2013.

Little, Roderick and Donald Rubin. 2002. Statistical Analysis With Missing Data. New York: Wiley.

Manski, Charles. 1989. “Anatomy of the Selection Problem.” *Journal of Human Resources* 24:343-360.

Meyer, Bruce, and James Sullivan. 2012. “Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure.” *Journal of Economic Perspectives*, 26(3), pp. 1-27.

Nicholas, Joyce and Michael Wiseman. 2010. “Elderly Poverty and Supplemental Security Income:2002-2005” *Social Security Bulletin*, 70(2), pp. 1-29.

Nicholas, Joyce and Michael Wiseman. 2009. “Elderly Poverty and Supplemental Security Income” *Social Security Bulletin*, 69(1), pp. 45-73.

NORC at the University of Chicago. 2011. “Assessment of the US Census Bureau’s Person Identification Validation System.” Final Report presented to the US Census Bureau.

Roemer, Mark. 2002. “Using Administrative Earnings Records to Assess Wage Data Quality in the Current Population Survey and the Survey of Income and Program Participation.” Longitudinal Employer-Household Dynamics Program Technical Paper No. TP-2002-22, US Census Bureau.

Rosenbaum, Paul and Rubin, Donald. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, 70(1), pp. 41-55.

Short, Kathleen. 2013. “The Research Supplemental Poverty Measure: 2012.” Current Population Reports, U.S. Census Bureau, P60-247.

Turek, Joan, Kendall Swenson, Bula Ghose, Fritz Scheuren, and Daniel Lee. 2012. “How Good Are ASEC Earnings Data? A Comparison to SSA Detailed Earnings Records” Presented at 2012 Federal Committee On Research Methodology Research Conference.

Venkatesh, Sudhir. 2006. *Off the Books: The Underground Economy of the Urban Poor*. Cambridge, MA: Harvard University Press.

Welniak, Edward J. 1990. “Effects of the March Current Population Survey’s New Processing System on Estimates of Income and Poverty.” US Census Bureau, Washington, DC, 1990.

Wooldridge, Jeffrey M. 2010. Econometric Analysis of Cross Section and Panel Data, Second Edition. Cambridge: MA, MIT Press.

Wooldridge, Jeffrey M. 2007. “Inverse Probability Weighted M-Estimation for General Missing Data Problems.” *Journal of Econometrics* 141: 1281-1301.

Ziliak, James P. 2006. “Understanding Poverty Rates and Gaps: Concepts, Trends, and Challenges.” *Foundations and Trends in Microeconomics* 1: 127–199.

Table 1: Assumptions Required for Consistent Estimation of Poverty for Alternative Estimators

Estimator	Missing at Random	Nonmatch at Random	ASEC correct	DER correct	Max(ASEC ,DER) correct	Independence of response/match/measurement conditional on poverty status
P^{C}	X		X			
P^{Full}			X			X
$P_{R,NR}^{Plug-in}$	X	X		X		X
$P_{NR}^{Plug-in}$		X	X	X		X
$P_{Max}^{Plug-in}$	X	X			X	X
P^{IPW}	X		X			
P^{LB}, P^{UB}			X			

Note: ASEC refers to the Annual Social and Economic Supplement of the Current Population Survey; DER refers to the Social Security Detailed Earnings Record; P^{C} is the official Census poverty measure based on the hot deck; P^{Full} is our full response poverty measure based on earnings in the DER and ASEC; $P_{R,NR}^{Plug-in}$ replaces ASEC earnings with DER earnings for all matched workers regardless of response status; $P_{NR}^{Plug-in}$ replaces ASEC earnings with DER earnings only for matched nonrespondents; $P_{Max}^{Plug-in}$ uses the maximum of the ASEC and DER for matched respondents; P^{IPW} uses inverse probability weights to reweight the ASEC poverty rate of respondents; and P^{LB}, P^{UB} are the Manski lower and upper bounds on the poverty rate. See the text for details.

Table 2: Summary Statistics (Head of Family)

	Respondent, DER Match		Respondent, DER Nonmatch		Nonrespondent, DER Match		Nonrespondent, DER Nonmatch	
Characteristic	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
Age	43.07	0.03	57.11	0.05	45.57	0.06	48.64	0.06
Gender								
Male (%)	55.70	0.12	49.13	0.12	55.17	0.22	54.99	0.18
Female (%)	44.30	0.12	50.87	0.12	44.83	0.22	45.01	0.18
Race								
White (%)	84.83	0.08	83.21	0.09	80.52	0.17	79.81	0.15
Black (%)	10.65	0.07	12.61	0.08	14.42	0.15	14.43	0.13
Other race (%)	4.52	0.05	4.18	0.05	5.06	0.09	5.76	0.08
Marital Status								
Married (%)	54.74	0.19	46.43	0.20	58.51	0.35	59.80	0.29
Widowed (%)	3.82	0.07	21.15	0.16	5.07	0.16	8.33	0.16
Separated or Divorced (%)	19.31	0.15	16.79	0.15	17.53	0.27	14.89	0.21
Single, Never-Married (%)	22.13	0.16	15.63	0.14	18.89	0.27	16.97	0.22
Educational Attainment								
Less Than High School (%)	9.13	0.07	23.28	0.10	11.65	0.14	15.57	0.13
High School Completed (%)	28.00	0.10	32.18	0.11	31.46	0.20	32.67	0.17
More than high school (%)	62.87	0.11	44.54	0.12	56.89	0.21	51.77	0.18
Employment Status								
Employed (%)	83.14	0.14	34.62	0.18	80.91	0.27	67.82	0.26
Unemployed (%)	3.78	0.07	2.65	0.06	3.22	0.12	2.68	0.09
Not in labor force								
Retired (%)	4.84	0.08	43.66	0.19	6.89	0.18	15.77	0.21
Disabled (%)	1.88	0.05	10.19	0.11	2.41	0.10	4.28	0.11
Other reason (%)	5.73	0.08	8.68	0.11	6.11	0.16	9.13	0.16
Family Size	2.51	0.00	2.13	0.00	2.69	0.01	2.69	0.01
Number of related children under 18	0.77	0.00	0.48	0.00	0.74	0.00	0.67	0.00

Official Poverty Status (%)	6.50	0.12	21.19	0.20	7.06	0.23	11.19	0.23
Family Type								
Married Couple (%)	53.47	0.11	44.49	0.11	57.07	0.20	57.93	0.16
Female Householder,no husband present (%)	26.74	0.10	36.87	0.11	25.41	0.17	25.34	0.14
Male Householder, no wife present (%)	19.79	0.09	18.64	0.09	17.52	0.15	16.74	0.12
ASEC Family Earnings ($)	58417	154	18164	128	62220	314	53614	271
DER Family Earnings ($)	55978	321	N/A	N/A	61974	1521	N/A	N/A
ASEC Family Income ($)	65504	162	39421	130	70124	328	63711	281
DER Family Income ($)	63065	327	N/A	N/A	69877	1527	N/A	N/A

Notes: Standard errors in parentheses are estimated using generalized function parameters. Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement.For information on sampling and nonsampling error, see <http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 3: Poverty Estimates Using ASEC and DER Data (Percent)

Year	P^{Full}	Std. Error	$P_{R,NR}^{Plug-in}$	Std. Error	$P_{NR}^{Plug-in}$	Std. Error	$P_{Max}^{Plug-in}$	Std. Error
1997	14.0	(0.216)	14.1	(0.217)	13.5	(0.213)***	12.8	(0.208)***
1998	13.7	(0.213)	13.5	(0.212)**	12.9	(0.208)***	12.3	(0.203)***
1999	12.5	(0.203)	12.6	(0.204)	12.1	(0.200)***	11.5	(0.196)***
2000	11.9	(0.198)	12.1	(0.199)**	11.6	(0.196)***	11.0	(0.191)***
2001	12.5	(0.143)	12.9	(0.145)***	12.2	(0.142)***	11.6	(0.139)***
2002	13.3	(0.146)	13.3	(0.146)	12.6	(0.143)***	12.0	(0.140)***
2003	13.3	(0.146)	13.5	(0.146)***	12.8	(0.143)***	12.3	(0.141)***
2004	13.8	(0.147)	13.8	(0.147)	13.2	(0.144)***	12.7	(0.142)***
2005	13.4	(0.145)	14.1	(0.148)***	13.2	(0.144)***	12.5	(0.140)***
2006	13.3	(0.143)	13.7	(0.145)***	13.0	(0.142)***	12.2	(0.138)***
2007	13.9	(0.145)	14.0	(0.146)	13.2	(0.142)***	12.5	(0.139)***
2008	14.2	(0.146)	14.5	(0.147)***	13.7	(0.144)***	12.9	(0.140)***

Notes: See the text and Table 1 for description of the poverty rates. Standard errors in parentheses are estimated using generalized function parameters. Significance reflects statistical test for comparison to P^{Full} ***p<0.01, **p<0.05, *p<0.10

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >.Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 4a: Components of Full Response Poverty Conditional on at Least One Earner (P^{Full}|Earner>=1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year	$P_R^{ASEC} * \Pr\{R\}$	$P_{NR,M}^{DER}$	$(P_{R,M}^{ASEC} - P_{R,M}^{DER})$	((2)+(3))* Pr{NR&M}	$(P_{R,NM}^{ASEC} - P_{R,M}^{ASEC})$	((2)+(3)+(5))* Pr{NR&DM}	(P^{Full}|Earner>=1)
1997	**6.8**	12.0	-1.4	**1.2**	2.9	**2.3**	**10.4**
1998	**6.4**	11.6	-1.4	**1.2**	2.7	**2.7**	**10.3**
1999	**6.2**	10.2	-1.4	**1.0**	2.9	**2.4**	**9.5**
2000	**5.2**	10.0	-1.4	**1.3**	2.5	**2.4**	**8.8**
2001	**5.5**	10.3	-1.8	**1.3**	3.1	**2.3**	**9.1**
2002	**5.5**	11.2	-1.9	**1.4**	3.9	**2.8**	**9.7**
2003	**5.5**	11.5	-2.1	**1.2**	1.6	**2.6**	**9.3**
2004	**5.5**	12.9	-1.8	**1.4**	1.3	**3.0**	**9.9**
2005	**5.5**	10.6	-2.0	**1.7**	7.2	**2.1**	**9.4**
2006	**5.6**	10.1	-1.8	**1.7**	6.7	**2.1**	**9.4**
2007	**5.8**	10.9	-1.8	**1.8**	8.0	**2.5**	**10.0**
2008	**6.2**	10.2	-1.7	**1.6**	8.1	**2.3**	**10.2**

Notes: See the text and Table 1 for description of the poverty rates. Bold columns sum to (P^{Full}|Earner>=1) subject to rounding error.

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 4b: Components of Full Response Poverty Conditional on No Earners (P^{Full}|Earner=0)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year	$P_R^{ASEC} * \Pr\{R\}$	$P_{NR,M}^{DER}$	$(P_{R,M}^{ASEC} - P_{R,M}^{DER})$	((2)+(3))* Pr{NR&M}	$(P_{R,NM}^{ASEC} - P_{R,M}^{ASEC})$	((2)+(3)+(5))* Pr{NR&DM}	(P^{Full}|Earner=0)
1997	**32.2**	0.0	0.0	**0.0**	35.8	**3.6**	**35.8**
1998	**30.8**	0.0	0.0	**0.0**	34.6	**3.8**	**34.6**
1999	**29.2**	0.0	0.0	**0.0**	32.3	**3.1**	**32.3**
2000	**29.2**	0.0	0.0	**0.0**	32.3	**3.1**	**32.3**
2001	**30.1**	0.0	0.0	**0.0**	33.9	**3.8**	**33.9**
2002	**31.3**	0.0	0.0	**0.0**	35.1	**3.8**	**35.1**
2003	**32.1**	0.0	0.0	**0.0**	36.2	**4.1**	**36.2**
2004	**32.2**	0.0	0.0	**0.0**	36.1	**3.9**	**36.1**
2005	**33.3**	0.0	0.0	**0.0**	36.9	**3.6**	**36.9**
2006	**32.7**	0.0	0.0	**0.0**	36.5	**3.8**	**36.5**
2007	**33.5**	0.0	0.0	**0.0**	36.8	**3.3**	**36.8**
2008	**33.4**	0.0	0.0	**0.0**	37.0	**3.6**	**37.0**

Notes: See the text and Table 1 for description of the poverty rates. Bold columns sum to (P^{Full}|Earner=0) subject to rounding error.

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 5: Poverty Estimates Using ASEC Data (Percent)

Year	$\boldsymbol{P^{Full}}$	Std. Error	$\boldsymbol{P_R^{ASEC}}$	Std. Error	$\boldsymbol{P^C}$	Std. Error	$\boldsymbol{P^{IPW}}$	Std. Error	Average ($\boldsymbol{P_R^{ASEC}}, \boldsymbol{P^{IPW}}$)	Std. Error
1997	14.0	(0.216)	14.2	(0.217)**	13.3	(0.211)***	13.5	(0.222)***	13.9	(0.333)
1998	13.7	(0.213)	13.9	(0.214)**	12.7	(0.206)***	13.2	(0.219)***	13.6	(0.329)
1999	12.5	(0.203)	13.0	(0.206)***	11.9	(0.199)***	12.3	(0.211)**	12.7	(0.316)
2000	11.9	(0.198)	12.4	(0.201)***	11.3	(0.193)***	11.6	(0.206)***	12.0	(0.309)
2001	12.5	(0.143)	13.1	(0.146)***	11.7	(0.139)***	12.3	(0.150)***	12.7	(0.225)*
2002	13.3	(0.146)	13.6	(0.147)***	12.1	(0.140)***	12.7	(0.151)***	13.2	(0.226)
2003	13.3	(0.146)	14.0	(0.149)***	12.5	(0.142)***	13.0	(0.152)***	13.5	(0.228)*
2004	13.8	(0.147)	14.1	(0.148)***	12.7	(0.142)***	13.2	(0.152)***	13.7	(0.228)
2005	13.4	(0.145)	13.8	(0.146)***	12.6	(0.141)***	13.1	(0.151)***	13.5	(0.225)
2006	13.3	(0.143)	13.8	(0.146)***	12.3	(0.139)***	13.0	(0.149)***	13.4	(0.224)
2007	13.9	(0.145)	14.0	(0.146)	12.5	(0.139)***	13.2	(0.150)***	13.6	(0.224)***
2008	14.2	(0.146)	14.5	(0.147)***	13.2	(0.142)***	13.8	(0.151)***	14.2	(0.227)

Notes: See the text and Table 1 for description of the poverty rates. Standard errors in parentheses are estimated using generalized function parameters. Significance reflects statistical test for comparison to P^{Full}. ***p<0.01, **p<0.05, *p<0.10

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 6: Full Response Poverty Rate and Official Poverty Rate for Select Demographic Groups

Full-Response Poverty (P^{Full})								
Year	Full Sample	Children	Adults	Elderly	Female Head	Black	White	Other Race
1997	14.0	21.1	11.7	10.6	31.9	26.6	11.9	15.9
1998	13.7	20.5	11.5	10.8	31.2	27.8	11.2	15.1
1999	12.5	18.2	10.9	9.8	28.2	24.0	10.4	15.6
2000	11.9	17.0	10.2	10.1	25.8	21.1	10.2	14.6
2001	12.5	17.7	11.0	10.2	27.4	23.3	10.8	13.4
2002	13.3	18.4	11.7	10.8	27.9	24.6	11.4	14.4
2003	13.3	18.6	11.8	10.2	28.4	24.9	11.3	16.0
2004	13.8	19.3	12.3	9.9	28.6	26.6	11.6	15.4
2005	13.4	18.9	11.9	10.3	30.0	26.6	11.3	14.5
2006	13.3	19.0	11.9	9.6	29.1	24.7	11.4	15.3
2007	13.9	20.1	12.2	9.9	29.4	27.0	11.8	13.5
2008	14.2	20.2	12.7	9.8	29.4	25.4	12.2	15.4
Official Poverty Rate (P^C)								
Year	Full Sample	Children	Adults	Elderly	Female Head	Black	White	Other Race
1997	13.3	19.9	10.9	10.5	31.5	26.5	11.0	16.1
1998	12.7	18.9	10.5	10.5	29.7	26.1	10.5	14.5
1999	11.9	17.1	10.1	9.7	27.5	23.6	9.8	14.5
2000	11.3	16.2	9.5	10.2	25.7	22.1	9.5	13.7
2001	11.7	16.3	10.1	10.1	26.5	22.7	9.9	12.8
2002	12.1	16.7	10.6	10.4	26.6	24.0	10.3	12.2
2003	12.5	17.6	10.8	10.2	27.5	24.3	10.6	13.5
2004	12.7	17.8	11.3	9.8	27.8	24.7	10.9	12.0
2005	12.6	17.6	11.1	10.1	28.7	24.8	10.7	13.0
2006	12.3	17.4	10.8	9.4	27.7	24.2	10.4	12.8
2007	12.5	18.0	10.9	9.7	27.7	24.5	10.6	12.1
2008	13.2	19.0	11.7	9.7	28.4	24.6	11.4	14.3

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 7: Population Weighted Demographic Full Response Poverty Rate

Year	P^{Full}	Std. Error	Population Weighted Demographic P^{Full}	Std. Error
1997	14.0	(0.216)	13.9	(0.215)
1998	13.7	(0.213)	13.6	(0.212)
1999	12.5	(0.203)	12.4	(0.203)
2000	11.9	(0.198)	11.8	(0.197)
2001	12.5	(0.143)	12.5	(0.143)
2002	13.3	(0.146)	13.2	(0.146)*
2003	13.3	(0.146)	13.1	(0.145)***
2004	13.8	(0.147)	13.7	(0.147)
2005	13.4	(0.145)	13.4	(0.145)
2006	13.3	(0.143)	13.2	(0.143)
2007	13.9	(0.145)	13.7	(0.145)**
2008	14.2	(0.146)	14.1	(0.146)

Notes: See the text and Table 1 for description of the poverty rates. Standard errors in parentheses are estimated using generalized function parameters.

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Appendix Table 1a: Nonresponse Probit Model (ASEC 1998-2002)

	ASEC 1998	ASEC 1999	ASEC 2000	ASEC 2001	ASEC 2002
Variable	Coeff. Est.	Coeff. Est.	Coeff. Est.	Coeff. Est.	Coeff. Est.
Metro Size (100,000-249,999)	0.071	-0.079	-0.105	-0.122	-0.068
	(0.030)*	(0.028)**	(0.028)**	(0.027)**	(0.021)**
Metro Size (250,000-499,999)	-0.03	-0.082	-0.118	-0.025	-0.047
	-0.026	(0.025)**	(0.025)**	-0.025	(0.020)*
Metro Size (500,000-999,999)	-0.05	-0.065	-0.104	-0.119	-0.072
	(0.024)*	(0.023)**	(0.022)**	(0.022)**	(0.018)**
Metro Size (1,000,000-2,499,999)	-0.079	-0.141	-0.063	-0.092	-0.095
	(0.020)**	(0.020)**	(0.020)**	(0.020)**	(0.015)**
Metro Size (2,500,000-4,999,999)	-0.055	-0.105	-0.077	-0.123	-0.123
	-0.029	(0.028)**	(0.028)**	(0.027)**	(0.022)**
Metro Size (5,000,000+)	-0.109	-0.089	-0.09	-0.158	-0.184
	(0.019)**	(0.019)**	(0.019)**	(0.019)**	(0.015)**
Midwest	0.134	0.145	0.124	0.078	0.053
	(0.020)**	(0.019)**	(0.019)**	(0.019)**	(0.015)**
South	0.108	0.147	0.107	0.051	0.057
	(0.019)**	(0.019)**	(0.019)**	(0.019)**	(0.015)**
West	0.232	0.231	0.187	0.222	0.153
	(0.019)**	(0.019)**	(0.019)**	(0.019)**	(0.015)**
Household Size	0.034	0.007	0.01	0.009	0.009
	(0.006)**	(0.006)	(0.006)	(0.006)	(0.004)*
Family Size	-0.211	-0.164	-0.164	-0.167	-0.139
	(0.013)**	(0.013)**	(0.012)**	(0.012)**	(0.009)**
Age	0.013	0.026	0.033	0.046	0.004
	(0.015)	(0.013)*	(0.014)*	(0.014)**	(0.013)
Age^2	-0.001	-0.001	-0.001	-0.002	0.000
	(0.000)	(0.000)**	(0.000)**	(0.000)**	(0.000)

Age^3	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)**	(0.000)**	(0.000)**	(0.000)
Age^4	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)**	(0.000)*	(0.000)**	(0.000)
Female	0.012	-0.001	0.009	0.024	0.005
	(0.015)	(0.015)	(0.015)	(0.014)	(0.011)
Black	-0.236	-0.153	-0.155	-0.189	-0.138
	(0.021)**	(0.021)**	(0.021)**	(0.020)**	(0.015)**
Native American	0.011	-0.144	0.037	-0.059	-0.012
	(0.061)	(0.057)*	(0.058)	(0.053)	(0.038)
Asian	-0.089	-0.167	-0.119	-0.182	-0.134
	(0.039)*	(0.037)**	(0.038)**	(0.035)**	(0.025)**
Married, Spouse Absent	0.089	-0.011	0.058	0.015	0.03
	(0.032)**	(0.030)	(0.031)	(0.031)	(0.025)
Previously Married	0.121	0.087	0.09	0.082	0.078
	(0.020)**	(0.019)**	(0.019)**	(0.019)**	(0.015)**
Never Married	0.073	0.001	0.055	0.053	0.013
	(0.023)**	(0.022)	(0.022)*	(0.022)*	(0.017)
Elementary School	0.105	0.117	0.132	0.063	0.093
	(0.028)**	(0.028)**	(0.029)**	(0.028)*	(0.024)**
Some High School	0.057	0.04	0.049	0.024	-0.014
	(0.023)*	(0.023)	(0.023)*	(0.023)	(0.018)
Some College	0.04	0.08	0.055	0.047	0.067
	(0.019)*	(0.018)**	(0.018)**	(0.018)**	(0.014)**
Associate's Degree	0.079	0.059	0.1	0.061	0.066
	(0.026)**	(0.025)*	(0.025)**	(0.025)*	(0.019)**
BA Degree	0.062	0.082	0.074	0.067	0.085
	(0.021)**	(0.020)**	(0.021)**	(0.020)**	(0.016)**
MS Degree	0.153	0.132	0.081	0.09	0.142
	(0.033)**	(0.031)**	(0.031)**	(0.031)**	(0.024)**

Professional Degree	-0.07	-0.071	-0.063	0.032	0.022
	(0.052)	(0.051)	(0.052)	(0.054)	(0.041)
PhD Degree	-0.08	-0.006	0.024	0.128	0.127
	(0.058)	(0.056)	(0.058)	(0.058)*	(0.046)**
Private Worker	-0.467	-0.682	-0.634	-0.657	-0.534
	(0.106)**	(0.101)**	(0.104)**	(0.106)**	(0.083)**
Federal Worker	-0.302	-0.55	-0.458	-0.542	-0.41
	(0.098)**	(0.092)**	(0.096)**	(0.099)**	(0.078)**
State Worker	-0.328	-0.593	-0.489	-0.561	-0.446
	(0.112)**	(0.106)**	(0.109)**	(0.111)**	(0.087)**
Local Government Worker	-0.402	-0.543	-0.526	-0.543	-0.472
	(0.109)**	(0.103)**	(0.107)**	(0.109)**	(0.085)**
Self-Employed	-0.92	-1.047	-1.036	-1.055	-0.923
	(0.107)**	(0.102)**	(0.106)**	(0.108)**	(0.085)**
Weeks worked last year	-0.001	-0.003	-0.002	-0.002	-0.003
	(0.001)	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Hours worked per week last year	0.000	0.001	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Foreign Born, U.S. Citizen	-0.05	-0.011	0.067	-0.039	-0.039
	(0.030)	(0.030)	(0.030)*	(0.028)	(0.023)
Foreign Born, Not a U.S. Citizen	0.026	0.03	0.036	-0.028	-0.049
	(0.028)	(0.026)	(0.026)	(0.025)	(0.020)*
Executive, Administrative, & Managerial	-0.095	0.159	0.1	-0.034	-0.027
	(0.100)	(0.093)	(0.098)	(0.099)	(0.078)
Professional Specialty	-0.038	0.2	0.1	0.003	-0.002
	(0.101)	(0.094)*	(0.098)	(0.100)	(0.079)
Technicians and related support	0.008	0.24	0.12	0.059	0.007
	(0.106)	(0.099)*	(0.103)	(0.105)	(0.083)
Sales	-0.142	0.074	-0.007	-0.088	-0.103
	(0.101)	(0.094)	(0.098)	(0.100)	(0.079)
Administrative Support	-0.094	0.177	0.061	-0.019	-0.061

	(0.100)	(0.093)	(0.097)	(0.099)	(0.078)
Private Household	-0.102	0.082	-0.097	-0.248	-0.166
	(0.133)	(0.131)	(0.126)	(0.130)	(0.106)
Protective Services	-0.081	0.231	0.114	-0.073	-0.079
	(0.110)	(0.104)*	(0.108)	(0.109)	(0.086)
Service	-0.119	0.115	-0.02	-0.042	-0.04
	(0.101)	(0.094)	(0.099)	(0.101)	(0.079)
Farming, Forestry, & Fishing	-0.031	0.147	0.015	-0.094	-0.141
	(0.106)	(0.100)	(0.103)	(0.105)	(0.083)
Precision Production: Craft and Repair	-0.051	0.185	0.065	0.007	0.000
	(0.100)	(0.094)*	(0.098)	(0.100)	(0.079)
Machine Operators, Assemblers, & Inspectors	-0.043	0.242	0.105	0.049	0.006
	(0.103)	(0.096)*	(0.101)	(0.103)	(0.081)
Transportation & Material Moving	-0.003	0.198	0.05	-0.019	-0.002
	(0.104)	(0.097)*	(0.101)	(0.103)	(0.081)
Handlers, Equipment Cleaners, Helpers, & Laborers	-0.131	0.065	-0.02	-0.041	-0.097
	(0.105)	(0.098)	(0.102)	(0.104)	(0.082)
Constant	1.54	1.359	1.35	1.207	1.56
	(0.162)**	(0.145)**	(0.156)**	(0.154)**	(0.135)**
Observations	55360	55941	56599	55013	86596

Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Appendix Table 1b: Nonresponse Probit Model (ASEC 2003-2009)

Variable	ASEC 2003 Coeff. Est.	ASEC 2004 Coeff. Est.	ASEC 2005 Coeff. Est.	ASEC 2006 Coeff. Est.	ASEC 2007 Coeff. Est.	ASEC 2008 Coeff. Est.	ASEC 2009 Coeff. Est.
Metro Size (100,000-249,999)	-0.012	-0.026	-0.015	-0.034	-0.026	-0.037	-0.002
	(0.021)	(0.022)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Metro Size (250,000-499,999)	-0.048	-0.064	-0.062	-0.1	-0.128	-0.038	-0.009
	(0.019)*	(0.020)**	(0.019)**	(0.019)**	(0.019)**	(0.020)	(0.020)
Metro Size (500,000-999,999)	-0.054	-0.053	-0.051	-0.043	-0.056	-0.025	0.009
	(0.018)**	(0.018)**	(0.017)**	(0.019)*	(0.019)**	(0.019)	(0.019)
Metro Size (1,000,000-2,499,999)	-0.058	-0.066	-0.105	-0.066	-0.097	-0.065	0.008
	(0.015)**	(0.015)**	(0.015)**	(0.016)**	(0.016)**	(0.016)**	(0.016)
Metro Size (2,500,000-4,999,999)	-0.134	-0.092	-0.099	-0.065	-0.057	-0.072	-0.055
	(0.022)**	(0.022)**	(0.017)**	(0.016)**	(0.016)**	(0.016)**	(0.016)**
Metro Size (5,000,000+)	-0.157	-0.206	-0.188	-0.192	-0.108	-0.116	-0.143
	(0.015)**	(0.015)**	(0.018)**	(0.017)**	(0.017)**	(0.017)**	(0.017)**
Midwest	0.043	0.049	0.146	0.1	0.121	0.088	0.064
	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**
South	0.026	0.006	0.064	0.081	0.144	0.102	0.041
	(0.015)	(0.015)	(0.014)**	(0.014)**	(0.014)**	(0.015)**	(0.014)**
West	0.108	0.123	0.177	0.206	0.243	0.194	0.12
	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**
Household Size	0.022	0.016	0.018	0.011	0.008	0.001	0.001
	(0.004)**	(0.004)**	(0.004)**	(0.004)*	(0.004)	(0.004)	(0.004)
Family Size	-0.159	-0.145	-0.174	-0.15	-0.171	-0.13	-0.176
	(0.009)**	(0.009)**	(0.009)**	(0.009)**	(0.010)**	(0.010)**	(0.010)**
Age	-0.007	0.01	0	-0.007	0.039	0.015	0.008
	(0.013)	(0.012)	(0.013)	(0.012)	(0.014)**	-0.014	(0.013)
Age^2	0.000	-0.001	0.000	0.000	-0.001	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)**	(0.000)	(0.000)

Age^3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)*	(0.000)	(0.000)
Age^4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)*	(0.000)	(0.000)	(0.000)	(0.000)*	(0.000)	(0.000)
Female	0.01	0.009	-0.016	0.003	-0.028	-0.007	-0.021
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)*	(0.012)	(0.012)
Black	-0.118	-0.157	-0.19	-0.16	-0.168	-0.182	-0.144
	(0.015)**	(0.015)**	(0.015)**	(0.016)**	(0.016)**	(0.015)**	(0.015)**
Native American	-0.008	-0.034	0.01	0.004	0.057	0.13	0.001
	(0.033)	(0.031)	(0.031)	(0.033)	(0.034)	(0.036)**	(0.033)
Asian	-0.123	-0.2	-0.227	-0.192	-0.147	-0.148	-0.141
	(0.025)**	(0.025)**	(0.024)**	(0.024)**	(0.024)**	(0.024)**	(0.024)**
Married, Spouse Absent	0.039	0.009	0.038	-0.015	0.084	0.028	0.049
	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)**	(0.025)	(0.025)
Previously Married	0.076	0.099	0.104	0.041	0.062	0.048	0.066
	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.016)**
Never Married	0.011	-0.028	0.007	-0.006	0.004	-0.027	0.023
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Elementary School	0.072	0.099	0.067	0.049	0.107	0.099	0.077
	(0.024)**	(0.024)**	(0.024)**	(0.025)*	(0.026)**	(0.027)**	(0.027)**
Some High School	0.015	0.058	0.037	0.008	0.088	0.044	0.05
	(0.018)	(0.019)**	(0.019)	(0.019)	(0.020)**	(0.020)*	(0.020)*
Some College	0.087	0.075	0.1	0.064	0.094	0.058	0.071
	(0.014)**	(0.014)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**
Associate's Degree	0.08	0.085	0.085	0.101	0.094	0.097	0.122
	(0.019)**	(0.019)**	(0.019)**	(0.019)**	(0.019)**	(0.019)**	(0.019)**
BA Degree	0.078	0.114	0.089	0.042	0.097	0.108	0.092
	(0.016)**	(0.016)**	(0.016)**	(0.016)**	(0.016)**	(0.016)**	(0.016)**
MS Degree	0.148	0.15	0.129	0.108	0.108	0.188	0.152
	(0.023)**	(0.023)**	(0.023)**	(0.024)**	(0.023)**	(0.023)**	(0.023)**

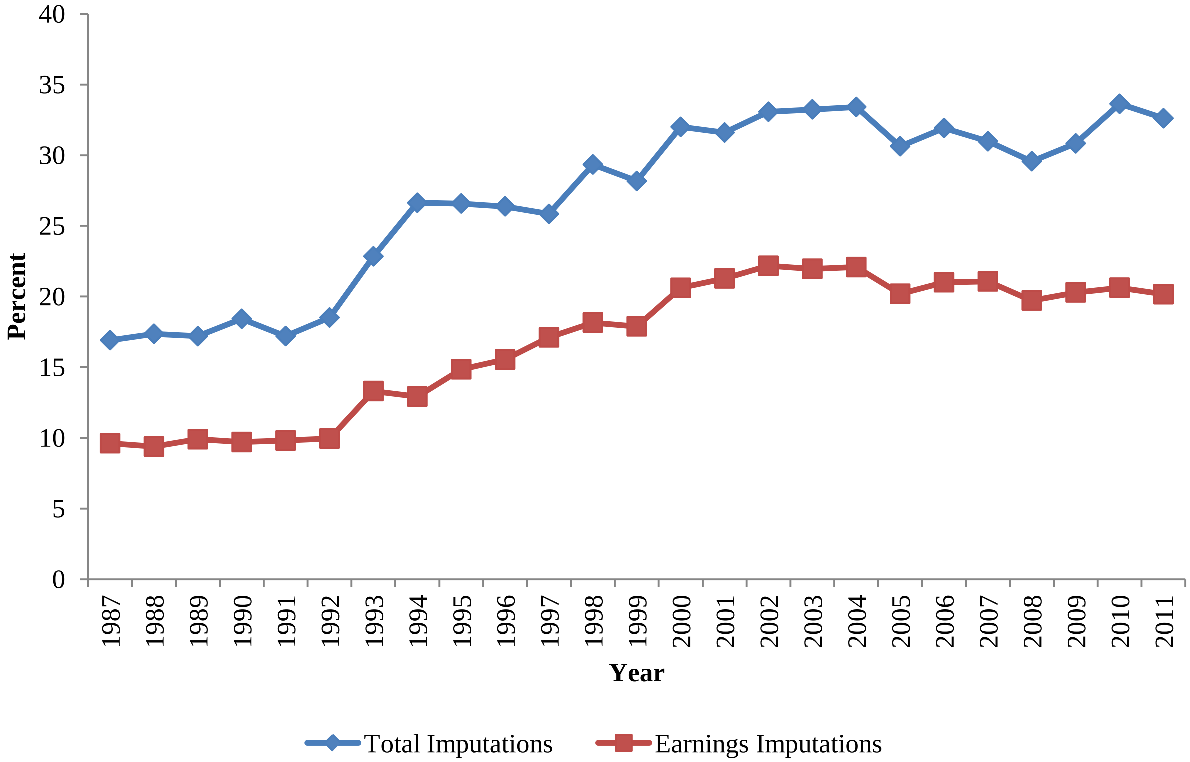
Professional Degree	0.055	0.093	-0.045	0.034	-0.012	0.034	0.026
	(0.044)	(0.044)*	(0.044)	(0.045)	(0.044)	(0.045)	(0.044)
PhD Degree	0.119	0.075	0.098	0.159	0.082	0.105	0.146
	(0.046)**	(0.045)	(0.046)*	(0.047)**	(0.046)	(0.046)*	(0.046)**
Private Worker	-0.703	-0.604	-0.568	-0.745	-0.739	-0.589	-0.686
	(0.082)**	(0.080)**	(0.087)**	(0.084)**	(0.087)**	(0.087)**	(0.085)**
Federal Worker	-0.589	-0.559	-0.483	-0.624	-0.598	-0.532	-0.634
	(0.076)**	(0.075)**	(0.081)**	(0.078)**	(0.081)**	(0.082)**	(0.079)**
State Worker	-0.582	-0.516	-0.469	-0.627	-0.604	-0.524	-0.602
	(0.086)**	(0.084)**	(0.091)**	(0.088)**	(0.091)**	(0.091)**	(0.089)**
Local Government Worker	-0.598	-0.516	-0.456	-0.612	-0.611	-0.514	-0.629
	(0.084)**	(0.083)**	(0.089)**	(0.087)**	(0.090)**	(0.090)**	(0.088)**
Self-Employed	-1.075	-0.973	-0.911	-1.067	-1.087	-0.973	-1.057
	(0.083)**	(0.082)**	(0.088)**	(0.086)**	(0.088)**	(0.088)**	(0.087)**
Weeks worked last year	-0.004	-0.003	-0.003	-0.003	-0.004	-0.004	-0.004
	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Hours worked per week last year	0.001	0.001	0.001	0.001	0.001	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)**	(0.001)
Foreign Born, U.S. Citizen	-0.015	0.015	-0.045	-0.023	-0.044	-0.039	-0.02
	(0.023)	(0.023)	(0.022)*	(0.023)	(0.022)	(0.022)	(0.022)
Foreign Born, Not a U.S. Citizen	0.024	0.001	-0.015	-0.051	-0.018	-0.001	0.013
	(0.020)	(0.020)	(0.020)	(0.020)*	(0.020)	(0.020)	(0.020)
Management	0.086	0.011	-0.13	0.119	0.162	-0.074	0.106
	(0.077)	(0.075)	(0.082)	(0.080)	(0.082)*	(0.082)	(0.080)
Business & Financial Operations	0.141	0.031	-0.041	0.186	0.168	-0.027	0.105
	(0.079)	(0.078)	(0.085)	(0.082)*	(0.085)*	(0.084)	(0.083)
Computer & Mathematical	0.206	0.152	-0.01	0.232	0.325	0.052	0.22
	(0.083)*	(0.082)	(0.088)	(0.086)**	(0.088)**	(0.087)	(0.086)*
Architecture & Engineering	0.135	0.093	0.042	0.197	0.2	-0.002	0.135
	(0.083)	(0.083)	(0.089)	(0.087)*	(0.089)*	(0.089)	(0.087)
Life, Physical, & Social Science	0.228	0.15	0.076	0.257	0.276	0.066	0.195

	(0.092)*	(0.091)	(0.098)	(0.095)**	(0.098)**	(0.098)	(0.097)*
Community & Social Services	0.1	-0.047	-0.03	0.182	0.202	-0.028	0.146
	(0.087)	(0.085)	(0.092)	(0.090)*	(0.091)*	(0.091)	(0.090)
Legal	0.041	0.011	-0.096	0.037	0.111	-0.077	0.198
	(0.094)	(0.091)	(0.097)	(0.096)	(0.098)	(0.097)	(0.096)*
Education, Training, & Library	0.133	0.077	-0.045	0.168	0.228	-0.036	0.115
	(0.080)	(0.078)	(0.085)	(0.083)*	(0.085)**	(0.085)	(0.083)
Arts, Design, Entertainment, Sports, & Media	0.073	0.03	-0.078	0.146	0.199	-0.047	0.092
	(0.084)	(0.083)	(0.090)	(0.088)	(0.090)*	(0.090)	(0.088)
Healthcare Pracitioner & Technical	0.081	0.025	-0.073	0.127	0.172	-0.019	0.067
	(0.080)	(0.078)	(0.085)	(0.083)	(0.085)*	(0.085)	(0.083)
Healthcare Support	0.034	-0.013	-0.103	0.157	0.153	-0.06	0.099
	(0.084)	(0.083)	(0.089)	(0.086)	(0.089)	(0.089)	(0.087)
Protective Services	-0.01	-0.004	-0.134	0.107	0.081	-0.061	0.086
	(0.084)	(0.082)	(0.089)	(0.087)	(0.088)	(0.089)	(0.087)
Food Preparation & Serving	0.024	-0.088	-0.212	0.027	0.083	-0.102	0
	(0.079)	(0.078)	(0.084)*	(0.082)	(0.085)	(0.085)	(0.083)
Building, Grounds Cleaning, & Maintenance	0.072	0.007	-0.076	0.111	0.123	-0.152	0.054
	(0.080)	(0.079)	(0.085)	(0.083)	(0.085)	(0.085)	(0.084)
Personal Care and Service	0.094	-0.066	-0.158	0.046	0.144	-0.037	0.062
	(0.081)	(0.080)	(0.086)	(0.084)	(0.087)	(0.086)	(0.084)
Sales and Related Occupations	0.01	-0.086	-0.182	0.033	0.058	-0.158	0.022
	(0.077)	(0.076)	(0.082)*	(0.080)	(0.082)	(0.082)	(0.081)
Office & Administrative Support	0.091	-0.001	-0.095	0.155	0.179	-0.035	0.116
	(0.076)	(0.075)	(0.082)	(0.079)	(0.082)*	(0.081)	(0.080)
Farming, Fishing, & Forestry	-0.01	-0.006	-0.082	0.078	0.19	0.012	0.159
	(0.095)	(0.093)	(0.099)	(0.099)	(0.102)	(0.102)	(0.101)
Construction Trades & Extraction Workers	0.048	0.015	-0.1	0.079	0.128	-0.083	0.028
	(0.078)	(0.077)	(0.083)	(0.081)	(0.083)	(0.083)	(0.082)

Installation, Maintenance, & Repair	0.059	0.008	-0.104	0.193	0.146	0.002	0.118
	(0.079)	(0.078)	(0.085)	(0.083)*	(0.085)	(0.085)	(0.084)
Production	0.183	0.024	-0.088	0.156	0.174	-0.018	0.183
	(0.078)*	(0.076)	(0.083)	(0.081)	(0.084)*	(0.083)	(0.082)*
Transportation & Material Moving	0.097	-0.027	-0.132	0.106	0.123	-0.094	0.113
	(0.078)	(0.077)	(0.084)	(0.081)	(0.084)	(0.083)	(0.082)
Constant	1.573	1.513	1.597	1.628	1.12	1.39	1.487
	(0.128)**	(0.128)**	(0.137)**	(0.123)**	(0.145)**	(0.148)**	(0.144)**
Observations	86613	85363	84682	84503	83543	84223	84741

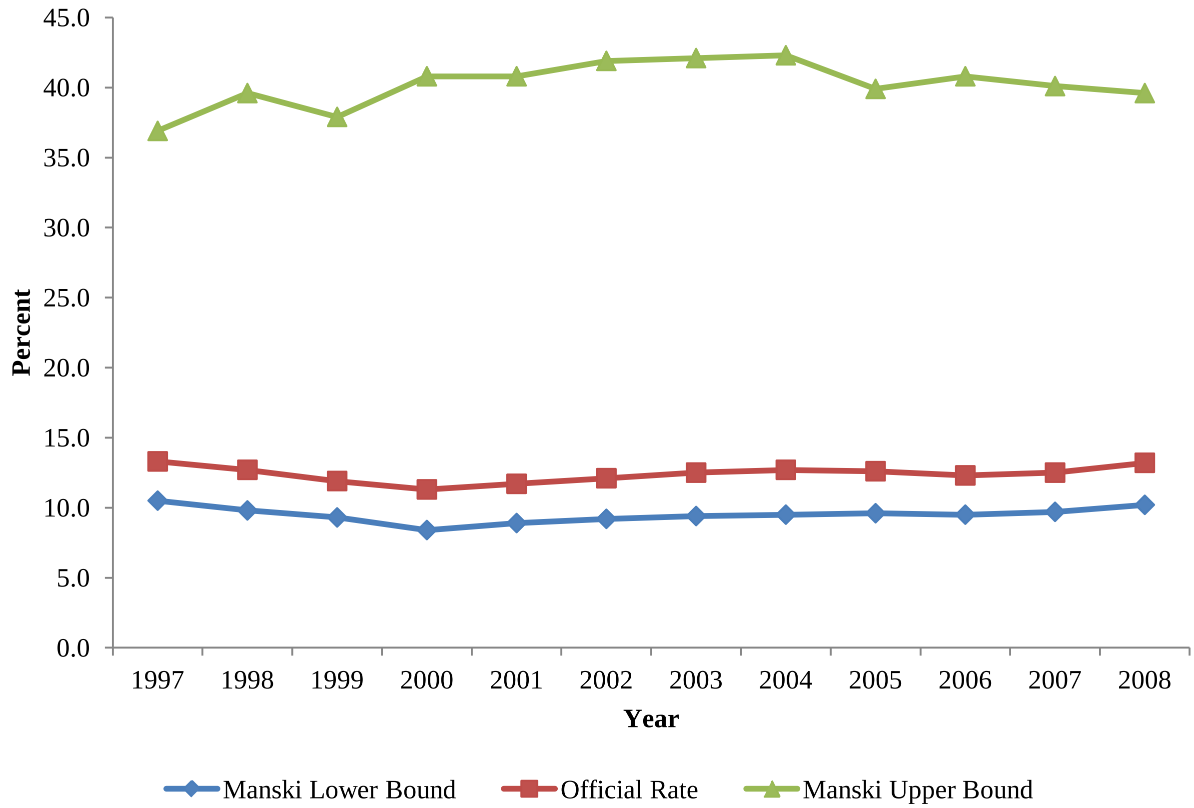
Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see < http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Figure 1: Trends in Earnings and Total (Item + Supplement) Imputations in the ASEC among Workers



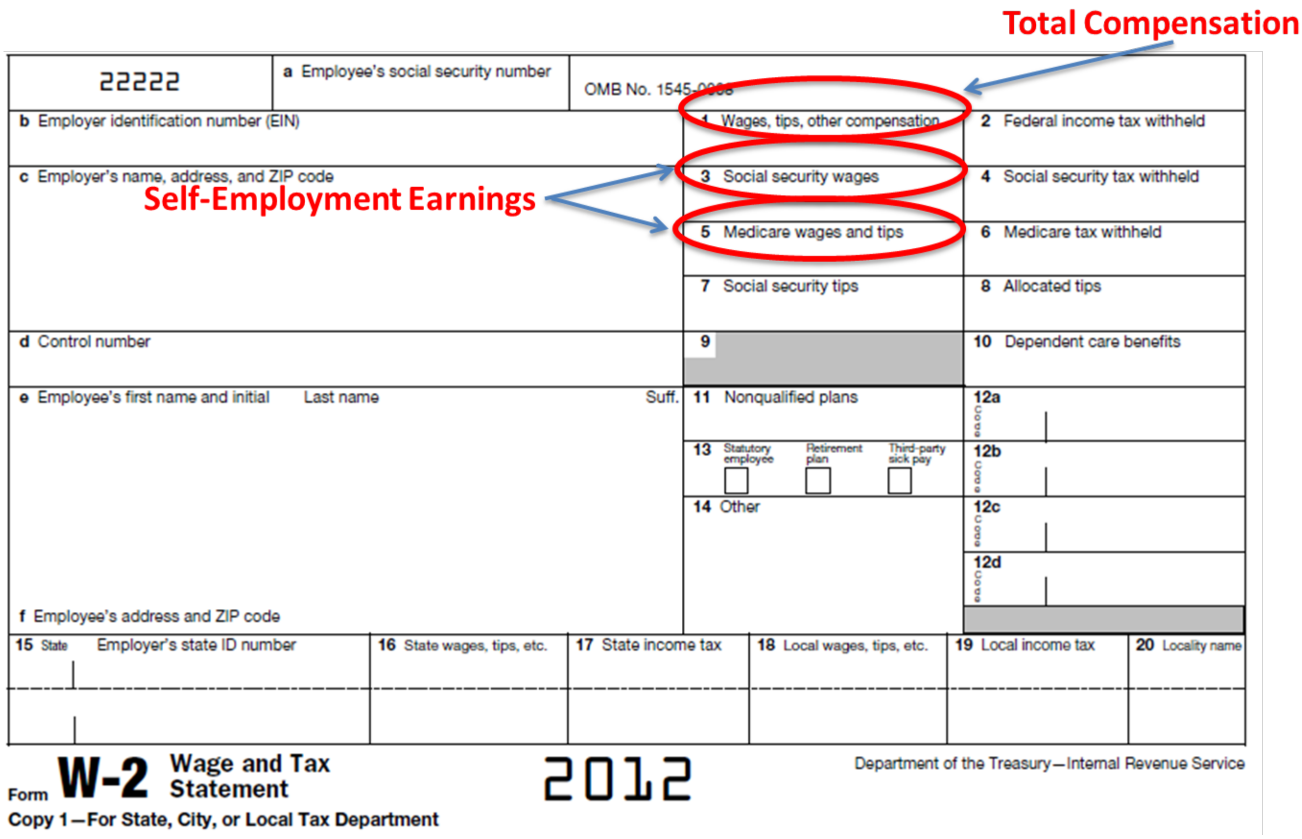
Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >.

Figure 2: Manski Bounds on the Official Poverty Rate



Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >.

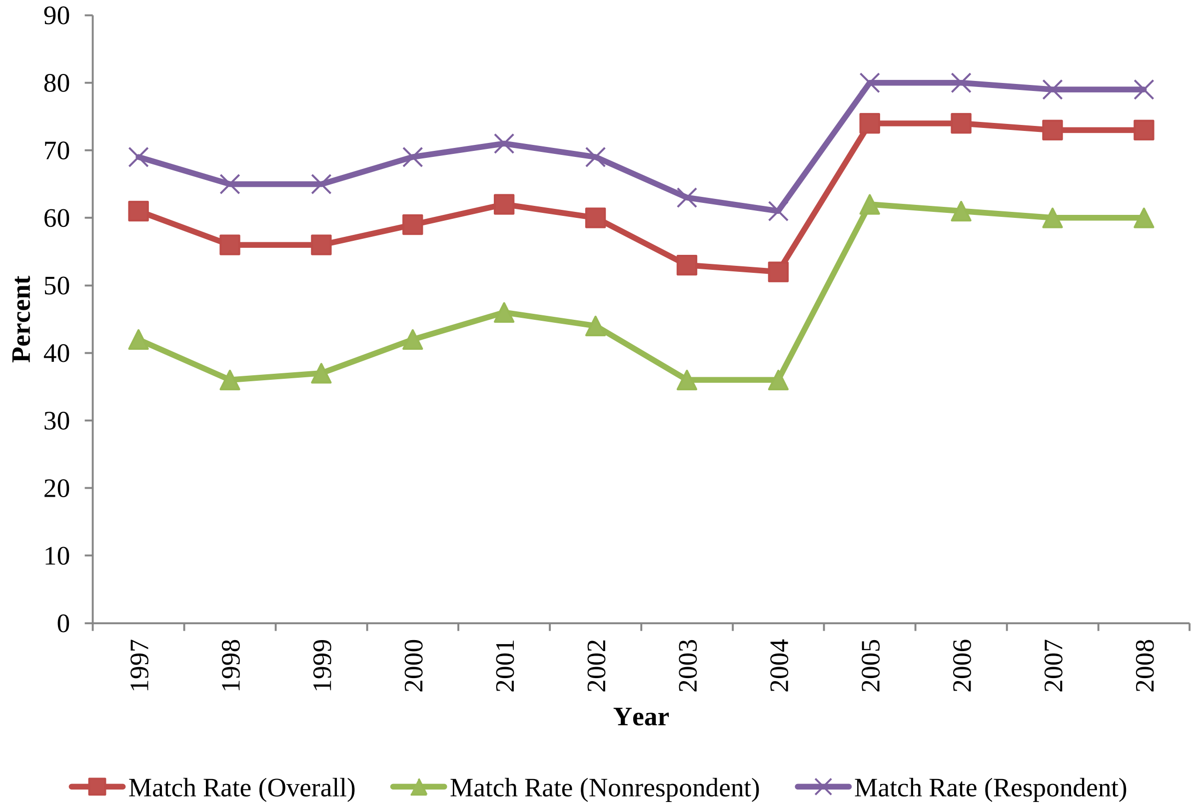
Figure 3: Sample W-2 Form



22222
a Employee's social security number
OMB No. 1545-0008
b Employer identification number (EIN)
1 Wages, tips, other compensation
2 Federal income tax withheld
c Employer's name, address, and ZIP code
3 Social security wages
4 Social security tax withheld
5 Medicare wages and tips
6 Medicare tax withheld
7 Social security tips
8 Allocated tips
d Control number
9
10 Dependent care benefits
e Employee's first name and initial Last name Suff.
11 Nonqualified plans
12a Code
13 Statutory employee Retirement plan Third-party sick pay
12b Code
14 Other
12c Code
12d Code
f Employee's address and ZIP code
15 State Employer's state ID number
16 State wages, tips, etc.
17 State income tax
18 Local wages, tips, etc.
19 Local income tax
20 Locality name

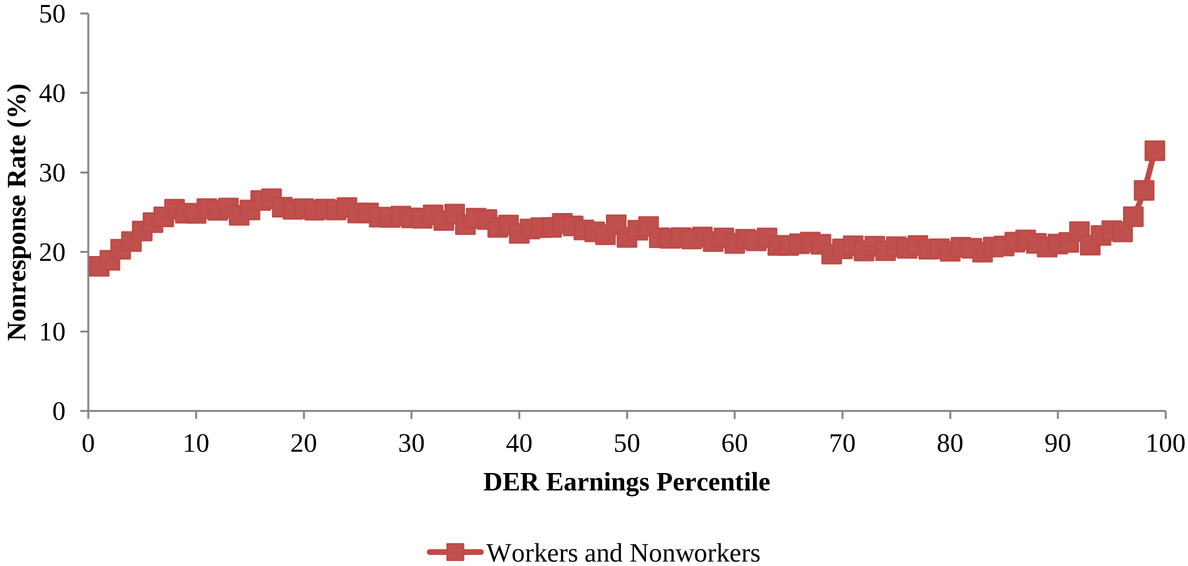
Form **W-2** Wage and Tax Statement **2012**
Department of the Treasury—Internal Revenue Service
Copy 1—For State, City, or Local Tax Department

Figure 4: Family Level ASEC-DER Match Rate (Earners)



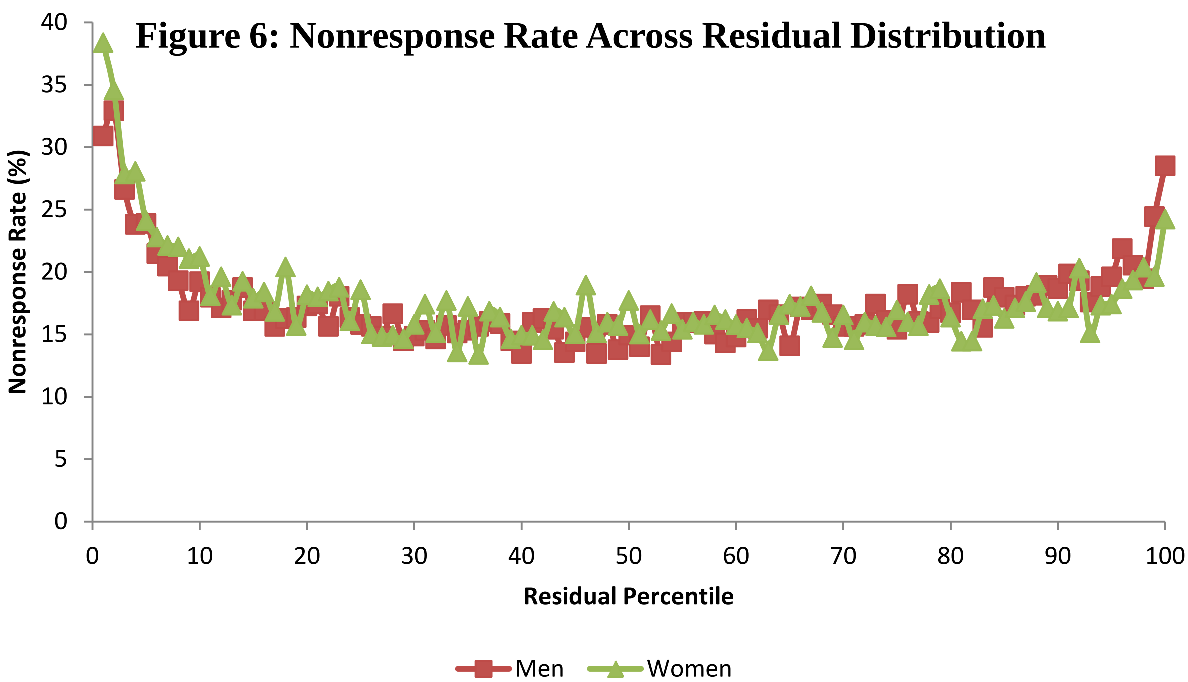
Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >. Social Security Administration, Detailed Earnings Record, 1997-2008.

Figure 5: Nonresponse Rate Across DER Earnings Distribution



Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >.

Figure 6: Nonresponse Rate Across Residual Distribution



Sources: Authors' calculations. U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <http://www.census.gov/prod/techdoc/cps/cpsmar13.pdf >.