Using Matched Household and Administrative Data to Measure Response Bias in Earnings

Christopher R. Bollinger[1], Barry T. Hirsch[2], Charles M.Hokayem[3], and James P. Ziliak[4]
[1]Department of Economics, University of Kentucky, Lexington, KY 40506
[2]Department of Economics, Georgia State University, Atlanta, GA 30302-3992
[3]Department of Economics, Centre College, Danville, KY 40422
[4]Department of Economics, University of Kentucky, Lexington, KY 40506

Abstract
Earnings non-response in household surveys is widespread, yet there is limited evidence on how response bias affects measured earnings. This paper examines the patterns and consequences of non-response using internal Current Population Survey worker records matched to administrative data on earnings for 2005-2010. Non-response across the earnings distribution, conditional on covariates, is found to be U-shaped for men and women, with left-tail "strugglers" and right-tail "stars" least likely to report earnings. Household surveys report too few low earners and too few very high earners. Non-response is ignorable over much of the distribution, but there exists trouble in the tails.

Key Words: CPS, hot deck imputation, non-response bias, earnings, measurement error

1. Introduction

Household surveys typically have high rates of earnings (and income) non-response. For example, the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the American Community Survey (ACS) have non-response rates on annual earnings of close to 20%. Individuals for whom earnings are not reported have their earnings "allocated" using hot deck imputation procedures that assign to them the earnings of a "similar" donor who has reported earnings. Despite the high rates of non-response to earnings questions in household surveys, we have limited knowledge regarding two important and closely related questions. First, is response bias ignorable; that is, do respondents and non-respondents have equivalent earnings, conditional on covariates?[1] Second, how do non-response and patterns of response bias vary across the earnings distribution among women and men?

In this paper, we address each of the questions using the March 2006-2011 CPS ASEC household files matched to administrative earnings records for calendar years 2005-2010.

---

[1] Following Rubin (1976) and Little and Rubin (2002), we use the term "missing at random" (MAR) to mean earnings data missing at random *conditional* on measured covariates. "Missing completely at random" (CMAR) refers to missingness (non-response) not dependent on earnings values, observable or not. Data are "not missing at random" (NMAR) if non-response depends on the value of missing earnings, conditional on covariates. The term "response bias" (or "non-ignorable response bias") is used as a synonym for NMAR.

We make substantial progress in addressing these fundamental questions. In what follows, we provide background on each issue, discuss the methods used to address them, describe the matched CPS-DER data, and present and interpret the evidence.

2. Background: Earnings Non-response, Match Bias, and Response Bias

Official government statistics, as well as most research analyzing earnings (and income) differences, include respondents and imputed earners in their analyses. Researchers implicitly assume that there is no systematic bias. This assumption is often unwarranted. For analyses of earnings differentials common in the social sciences, inclusion of workers with imputed earnings can cause a large systematic bias in earnings gap estimates with respect to wage determinants that are not imputation match criteria or are imperfectly matched in the hot deck procedure. This "match bias" (Hirsch and Schumacher 2004; Bollinger and Hirsch 2006) occurs even if non-response is missing completely at random.

Although match bias can be substantial, it is easy to (largely) eliminate. Among the remedies are: exclude imputed earners from the analysis; exclude the imputations and reweight the sample by the inverse probability of response; retain the full sample but adjust estimates using a complex correction formula; or retain the full sample but conduct one's own earnings imputation procedure using all earnings covariates in one's model. In practice, each of these approaches eliminates first-order match bias and produces highly comparable results (Bollinger and Hirsch 2006). Each of these methods, however, assumes earnings are missing at random (MAR), thus assuming response bias is ignorable.[2]

The validity of the MAR assumption is difficult to test with public use data. A direct approach for determining whether non-response is ignorable is to conduct a validation survey in which one compares CPS household earnings data with administrative data on earnings provided for both CPS earnings respondents and non-respondents. That is the approach used here. We are not the first study to examine response bias in this way, but prior studies examining CPS non-response are dated, use small samples, and examine restricted populations (e.g., married white males).

Most similar to our analysis is a paper by Greenlees et al. (1982), who examine the March 1973 CPS and compare reported wage and salary earnings in 1972 to matched income tax records. They restrict their analysis to full-time, full-year male heads of households in the private nonagricultural sector whose spouse did not work. They conclude that non-response is not ignorable, with response negatively related to earnings (negative selection into response). The imputations understate administrative earnings of the non-respondents by 0.08 log points. Herriot and Spiers (1975) earlier reported similar results using these data, the ratio of CPS respondent to IRS earnings being 0.98 and of CPS imputed to IRS earnings being 0.91.

It is not known whether results from these early studies can be generalized outside their time period and demographic samples. There has been limited study of CPS response bias using recent data, and these studies have not examined differences in non-response across the distribution. Given the increase in non-response over time, it is important to know whether non-response is ignorable and, if not, the size and patterns of bias.

[2] Although inclusion of imputed earners in the estimation sample can introduce severe match bias, it does *not* correct for response bias since donor earnings assigned to non-respondents are drawn from the sample of respondents. Earnings of non-respondents are not observed.

3. The CPS ASEC Imputation Procedure for Earnings

The Census Bureau has used a hot deck procedure for imputing missing income since 1962. The current system has been in place with few changes since 1989 (Welniak 1990).[3] The CPS ASEC uses a sequential hot deck procedure to address item non-response for missing earnings data. This procedure assigns individuals with missing earnings values that come from individuals ("donors") with similar characteristics. The ASEC sequential hot deck procedure for earnings variables first divides individuals with missing data into one of 12 allocation groups defined by the pattern of non-response (e.g., only missing earnings from longest job, or missing both longest job information and earnings). Second, an observation in each allocation group is matched to a donor with complete data based on a large set of socioeconomic match variables. If no match is found based on the large set of variables, then a match variable is dropped and variable definitions collapsed (i.e., categories are broadened) to be less restrictive. This process is repeated until a match is found. When a match is found, the missing earnings amount is replaced with the reported earnings from the first available matched donor.

The sequential hot deck used in the CPS ASEC has the advantage that it always finds a match in the current month. Disadvantages are that one cannot know which attributes are matched or the extent to which variables were collapsed. The quality of an earnings match depends on how common are an individual's attributes (Lillard et al. 1986).

The CPS ASEC also includes "whole imputes." Whole imputation refers to households who participated in the monthly CPS, but refused participation in the ASEC supplement. The entire supplement is replaced (imputed) by a "similar" household participating in the supplement. Whole imputes account for about 10% of ASEC records. Households who did not participate in the ASEC supplement have their earnings included in the matched administrative earnings data described below. We do not directly observe their household characteristics because it is the donor household that is included in the CPS. For this reason, whole imputes are excluded from our analysis.

4. Data Description: The CPS-DER Earnings Match Files

The data used in our analysis are Current Population Survey (CPS) person records matched to administrative earnings records. We use Census internal CPS ASEC files for survey years 2006-2011 reporting earnings for calendar years 2005-2010. The internal files have top-coded values for income sources substantially higher than the public use top codes.[4] CPS files are matched to the SSA's Detailed Earnings Record (DER) file. The DER file is an extract of SSA'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. Only positive self-employment earnings are reported in DER (Nicholas and Wiseman 2009). The DER file contains all earnings reported on a worker's W-2 forms. These earnings are not capped at the FICA contribution amounts and include earnings not covered by Old Age Survivor's Disability Insurance (OASDI) but subject to the Medicare

[3] The sequential hot deck procedures used in the March survey prior to 1989 were fairly primitive, with schooling not a match variable until 1975. Lillard, Smith, and Welch (1986) provided an influential critique of Census methods. Welniak (1990) documents changes over time in Census hot deck methods for the March CPS.

[4] Larrimore et al. (2008) designate differences in top code values for the internal and public files.

tax. It is helpful that DER earnings are not capped given the concerns regarding non-response and response bias in the right tail of the distribution. We cap DER annual earnings at $2 million to avoid influence from extreme earnings on estimated wage equation coefficients. Our $2 million cap "roughly matches" the cap on annual earnings in the internal CPS ASEC files.[5]

The DER file contains deferred wages such as contributions to 401(k), 403(b), 408(k), 457(b), 501(c), and HSA plans. It does not include some components of compensation, for example, pre-tax health insurance premiums and education benefits (Abowd and Stinson 2013). The DER file cannot measure earnings that are off the books and not reported to IRS and SSA. We compare differences in CPS earnings (likely to include undocumented earnings) and DER earnings for samples with and without demographic or occupational groups of workers most likely to have undocumented earnings.

The Center for Administrative Records Research and Applications (CARRA) at the Census Bureau matches the DER file to the CPS 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.[6] This assignment relies on a probabilistic matching model based on name, address, date of birth, and gender. CPS workers not matched to DER are disproportionately low wage workers and in occupations where off-the-books earnings are most common. Bond et al. (2013) provide similar evidence using administrative data matched to the American Community Survey (ACS).

Because workers can appear multiple times each 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 earnings (Box 1 of W-2, labeled "Wages, tips, other compensation") across all employers. In this way, DER earnings is most compatible with CPS earnings from all wage and salary jobs (WSAL-VAL). We classify a worker as having imputed earnings if either wages and salary from the longest job (I-ERNVAL) or from other jobs (I-WSVAL) is imputed. We construct CPS and DER average hourly wages by dividing annual CPS or DER earnings by annual hours worked. Annual hours worked comes from multiplying weeks worked (WKSWORK) by usual hours worked per week (HRSWK).

Match rates between the CPS and DER administrative data among earners beginning with the 2006 ASEC are about 85 percent. The regression sample used in our analysis includes full-time, full-year, non-student wage and salary workers ages 18 to 65 who have positive CPS and DER earnings reported for the prior calendar year. This 2006-2011 CPS-DER regression sample includes 287,704 earners, 157,041 men and 130,663 women. Earnings non-response rates among this sample is 19.5% among men and 19.3% among women.

Table 1 provides summary statistics for our sample by gender. We focus on measures of earnings and earnings response. For men, overall weighted mean earnings in the CPS and in DER are roughly equivalent; among women CPS earnings are moderately higher. However, using the mean of log wages, CPS earnings exceed DER earnings by about 7% for men and women (0.067 and 0.071 log points). The seeming inconsistency arises from

---

[5] The two components of our CPS total earnings variable, earnings on the primary job and all other earnings, are each capped at $1.1 million.

[6] Prior to the 2006 ASEC (calendar year 2005), the CPS collected respondents' SSN and an affirmative "opt-in" agreement allowing a match to administrative data. In 2006, Census switched to an "opt-out" option. Prior to this change, match rates among earners were considerably lower.

exponentiation of log differences can substantially overstate the arithmetic percentage difference if focal earnings (the CPS) has lower dispersion than reference earnings (DER) (Blackburn 2007). Very high earnings are far more common among men than women.

Table 1: Selected Summary Statistics for Estimation Sample

	Men		Women		
Characteristic	Mean	Std. Dev.	Mean	Std. Dev.	Difference
CPS ASEC Wage ($2010)					
Full Sample	$27.05	$27.14	$20.80	$18.57	$6.26
CPS Respondents	$27.11	$26.16	$20.94	$18.37	$6.18
CPS Non-respondents	$26.81	$30.83	$20.22	$19.38	$6.59
lnW–CPS)	3.075	0.652	2.849	0.604	0.23
DER Wage ($2010)					
Full Sample	$27.44	$63.63	$19.87	$17.33	$7.57
CPS Respondent	$27.05	$54.51	$20.01	$17.03	$7.04
CPS Non-respondent	$29.06	$92.11	$19.31	$18.53	$9.75
lnW–DER)	3.008	0.782	2.778	0.691	0.23
Non-response Rate (%)	19.5	39.6	19.3	39.5	0.23
Observations	157,041		130,663		

Note: All means are weighted using CPS ASEC Supplement weights. Sources: U.S. Census Bureau, Current Population Survey, 2006-2011 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 2005-2010.

For responding men, DER wages ($27.05 in 2010$) are not statistically different from CPS wages for these same men ($27.11), but for responding women DER wages ($20.01) are lower than their CPS wages ($20.94). For non-responding men, their imputed CPS wages ($26. 81) are substantially lower than their DER wages ($29.06). The opposite pattern is seen among non-responding women, whose imputed CPS hourly earnings is an average $20.22, as compared to their $19.31 DER wage. Focusing just on DER wages, CPS male non-respondents exhibit higher DER wages than do respondents ($29.06 versus $27.05), whereas among women non-respondents exhibit lower DER wages than do respondents ($19.31 versus $20.01).

5. Is Response a Function of Earnings? Non-Response across the Distribution

Although evidence is limited, previous studies have concluded that there is negative selection into response. That is, as true earnings rise, so does non-response. We initially follow the approach by Greenlees et al. (1982), who measure the likelihood of CPS response as a function of administrative (i.e., DER) earnings matched to the CPS, conditional on a rich set of covariates. The Greenlees et al. analysis was conducted for white males working full-time/full-year married to non-working spouses.

To explore the relationship between non-response and earnings, the following model of non-response using our matched CPS-DER sample is estimated:

$$NR_i = \theta \, lnW - DER_{\,i} + X_i\beta + u_i \qquad (1)$$

where NR_i represents individual i's earnings non-response status (0 or 1),W−DER is the DER wage, and X_i includes a detailed set of controls (potential experience, race, marital status, citizenship, education, metropolitan area size, occupation, industry, and year). We then move from use of a single linear log wage term to categorical measures for all wage percentiles, thus allowing different responses throughout the earnings distribution.

$$NR_i = \theta_k DER\ Wage\ Percentile_{ik} + X_i\beta + u_i \qquad (2)$$

Table 2 provides estimates of the non-response to earnings relationship using linear probability models, with and without a detailed set of controls, along with the corresponding marginal effects (evaluated at the means) using probit estimation. OLS results are highly similar to those from probit. We first examine θ, the coefficient on ln*Wage*, as in Greenlees et al., which measures the central tendency of non-response with respect to the wage. The top panel of Table 2 provides results for men and the middle panel for women. Full results are available from the authors.

Table 2: CPS Mean Non-response with Respect to DER Wages for Men, Women, and 'Mad Men', 2006-2011

	OLS	Probit Marginal Effects	OLS w/X's	Probit w/X's Marginal Effects
		Men		
lnW–DER	-0.0178***	-0.0164***	-0.0124***	-0.0107***
	(0.00149)	(0.00140)	(0.00179)	(0.00163)
Constant	0.234***		0.291***	
	(0.00458)		(0.0128)	
Observations	157,041	157,041	157,041	157,041
R-squared	0.001		0.019	
		Women		
lnW–DER	-0.0357***	-0.0332***	-0.0397***	-0.0356***
	(0.00177)	(0.00161)	(0.00221)	(0.00192)
Constant	0.282***		0.307***	
	(0.00501)		(0.0147)	
Observations	130,663	130,663	130,663	130,663
R-squared	0.004		0.020	
		Mad Men †		
lnW–DER	0.0178***	0.0160***	0.0134***	0.0114***
	(0.00210)	(0.00184)	(0.00248)	(0.00210)
Constant	0.106***		0.139***	
	(0.00673)		(0.0299)	
Observations	78,179	78,179	78,179	78,179
R-squared	0.001		0.019	

*** p<0.01, ** p<0.05, * p<0.1. †'Mad Men' sample includes married, white, male U.S. citizens with spouse present. For sources, see the note in Table 1.

In contrast to Greenlees et al. (and other prior literature), our coefficients on earnings in Table 2 are negative rather than positive for both men and women. This suggests a central tendency of positive rather than negative selection into response. That said, the OLS coefficient for men (with controls) is close to zero (-0.012 with s.e. 0.002), although highly significant. Among women, we obtain a larger coefficient (-0.040 with s.e. 0.002), again indicating that on average non-response declines with earnings, conditional on covariates.

Although these results provide what we believe are accurate measures of central tendency for these broad samples of men and women, such results are not particularly informative. Our concerns are two-fold. First, the Greenlees et al. result showing the opposite central tendency from that seen in Table 2 was for a small 1972 sample not representative of today's workforce. Second, the relationship between non-response and earnings may vary over the distribution, making measures of central tendency misleading. Non-response may decline, remain constant, or increase with respect to earnings over different ranges of the distribution, a possibility not examined in prior studies.

To compare our results with Greenlees et al., we restrict our sample to married white men who are citizens, with spouse present. Unlike Greenlees et al., we include those with working spouses since married women's labor force participation is now closer to the norm than the exception. We refer to this as a "Mad Men" sample, shown in the bottom panel of Table 2. This sample is likely to have few workers in the left tail of the DER distribution. In contrast to the negative earnings coefficients of -0.018 and -0.012 for all full-time/full-year men (cols. 1 and 3), using the Mad Men sample flips the signs and produces coefficients of 0.018 and 0.013 (each with s.e. 0.002), consistent with Greenlees et al. and prior studies finding negative selection into response.

Rather than focus on central tendency, it is far more informative to examine how non-response varies across the distribution. The well-known paper by Lillard et al. (1986, p. 492) speculated that CPS non-response is likely to be highest in the tails of the distribution (U-shaped), but to the best of our knowledge, no study has directly provided such evidence. Since we cannot observe reported CPS earnings for non-respondents, it is difficult to examine this relationship absent matched administrative data on earnings.

Patterns of non-response across the entire distribution are most easily discerned visually. In Figure 1, we show non-response rates for both men and women for each percentile of the DER wage distribution. The top curve for each shows the unadjusted mean rate of non-response at each percentile of the DER wage distribution. The lower curve for each is based on equation (2), which includes a large set of covariates and a full set of percentile dummies (with one omitted percentile). We follow Suits (1984) and adjust the values of all the percentile dummy coefficients (along with the "zero" omitted percentile) to provide a measure of the conditional non-response rate at each percentile, relative to the mean rate.[7] By construction, the 100 values shown in the lower curve sum to zero.

In the top half of Figure 1 we show male non-response rates for each percentile of the DER wage. The pattern here is U-shaped, with considerably higher non-response in the lower and upper tails of the distribution, but with rather constant non-response rates from about the 20th to 95th percentiles. There is little difference between the unadjusted (top) and adjusted (bottom) curves, apart from the downward adjustment of the latter curve to reflect measurement relative to the conditional mean rate. Whereas we see non-response decline in the left tail throughout much of the first quintile, rising non-response is restricted to the top ventile. Non-response is largely constant throughout the wage distribution, the obvious exceptions being in the tails.

[7] The Suits (1984, p. 178) adjustment factor is the value k that makes the average of the percentile coefficients equal to zero. That is, $k = -(b_2 + b_3 + \ldots b_{100} + 0)/100$, where b represents the 99 included percentile dummies. The value k is added to each b and to "zero" for the omitted percentile. These Suits-adjusted coefficients are shown in the lower curves in Figure 1.

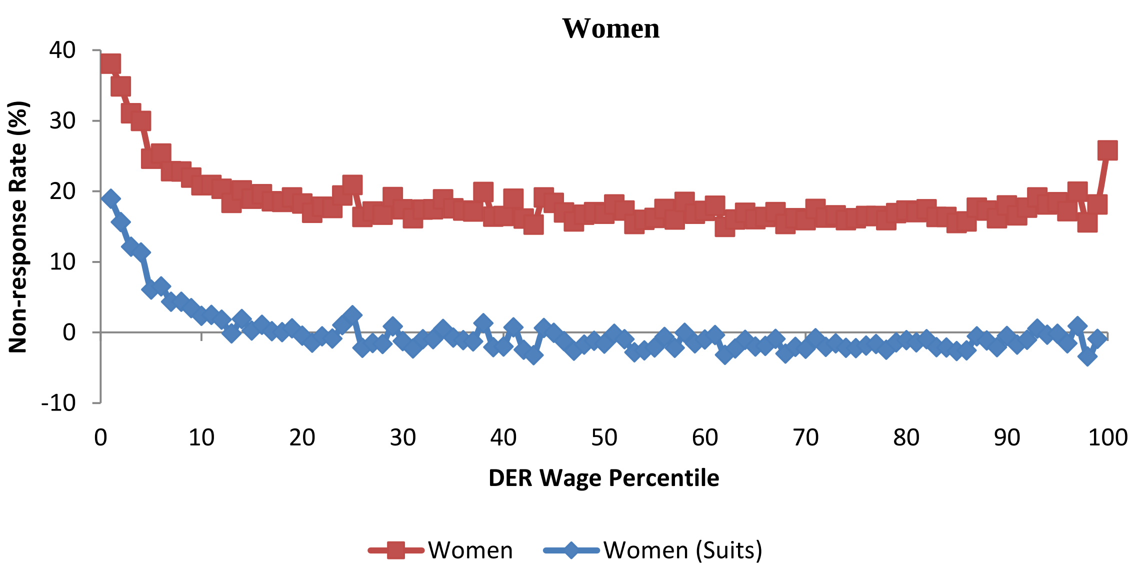


Figure 1: Earnings Non-response Rates and Conditional Response Rates Relative to Mean by Percentiles over the Male and Female DER Wage Distributions

Squares show unadjusted non-response rates. Diamonds show rates adjusted for covariates, relative to the mean. See text for details. For sources, see the note in Table 1.

The evidence for women (lower half of Figure 1) is qualitatively the same as for men, with a U-shaped non-response pattern. That said, there are differences in the magnitudes of the tails. In the lower-end of the wage distribution, women exhibit higher rates of adjusted and unadjusted non-response than do men. High non-response for earnings (and other income sources) among low-wage women may result in part from the (invalid) concern that reporting such information to Census might place income support program eligibility at risk. In the right tail, women exhibit minimal increases in non-response until one moves to the highest percentile. Rather than characterizing this pattern as "U-shaped" emphasis should be given to the high female rates of non-response in the left tail coupled with similar rates throughout the rest of the distribution outside of the top percentile.[8]

The male and female non-response curves shown across the wage distribution in Figure 1 are based on gender-specific wage percentiles. At a given percentile, the wage for men will be considerably higher than that for women. In a figure not shown (available on request), we form percentiles based on the joint male-female DER wage distribution and show the unadjusted non-response rates for men and women at each percentile of this common distribution. The male and female curves are remarkably similar, indicating that women and men have similar likelihood of non-response at similar wage levels. The patterns evident in Figure 1 result in part from women being highly concentrated in the left tail and men in the right tail. Based on the joint earnings distribution, male and female non-response behaviors are similar when compared at the same wage levels.

Our interpretation of the evidence is straightforward. The good news is that earnings non-response in the CPS appears to be largely ignorable throughout much of the earnings distribution, varying little with the realized level of earnings, conditional on covariates. To the extent that there is a pattern over the 20th to 95th percentiles, it is one consistent with weak positive selection into response, with non-response declining slightly over much of the distribution before turning up at very high levels of earnings. Where there most clearly exist problems is in the tails. Non-response is highest among "strugglers" and "stars". Characterizing selection into response based solely on estimates of central tendency over entire distributions is largely uninformative and potentially misleading.

Rates of non-response are particularly high in the lower decile of wage distributions. There are substantial disparities between CPS and DER earnings in the left tail; some of this difference being the result of off-the-books earnings, which we briefly examine below. In the right tail, high non-response is seen primarily in the highest two percentiles for men and the top percentile for women. These percentiles correspond roughly to where individual earnings are top coded in public use CPS files. Analysis of workers with top-coded earnings is already difficult for researchers using public use files; high non-response among such earners makes such research all the more difficult.[9]

---

[8] Coefficients on control variables in the non-response equations (available on request) provide information on which types of workers are least and most likely to not respond to the ASEC earnings questions, conditional on the wage (using the full set of percentile dummies). For the most part, demographic, location, and job-related measures account for little of the variation in response. Coefficients are generally similar for men and women. Most notable are high non-response probabilities found among workers who are black, Asian, never married, and residents in large (5 million plus) metro areas. Public sector workers are more likely to report earnings.

[9] Researchers using the CPS often assign mean earnings above the top-code based on information provided by Census or by researchers using internal CPS files (Larrimore et al. 2008). Because very high earners are less likely to report earnings in the CPS, there will be some understatement of high-end earnings due to non-ignorable response bias.

6. Further Evidence on Response Bias: Residuals across the Distribution

In the previous section, we provided evidence of response bias based on rates of non-response across the DER wage distribution, conditional on covariates. An alternative way to exhibit the same pattern is to examine differences in wage residuals across the distribution for CPS respondents and non-respondents, with residuals drawn from DER wage equations in which administrative earnings data are observed for both.

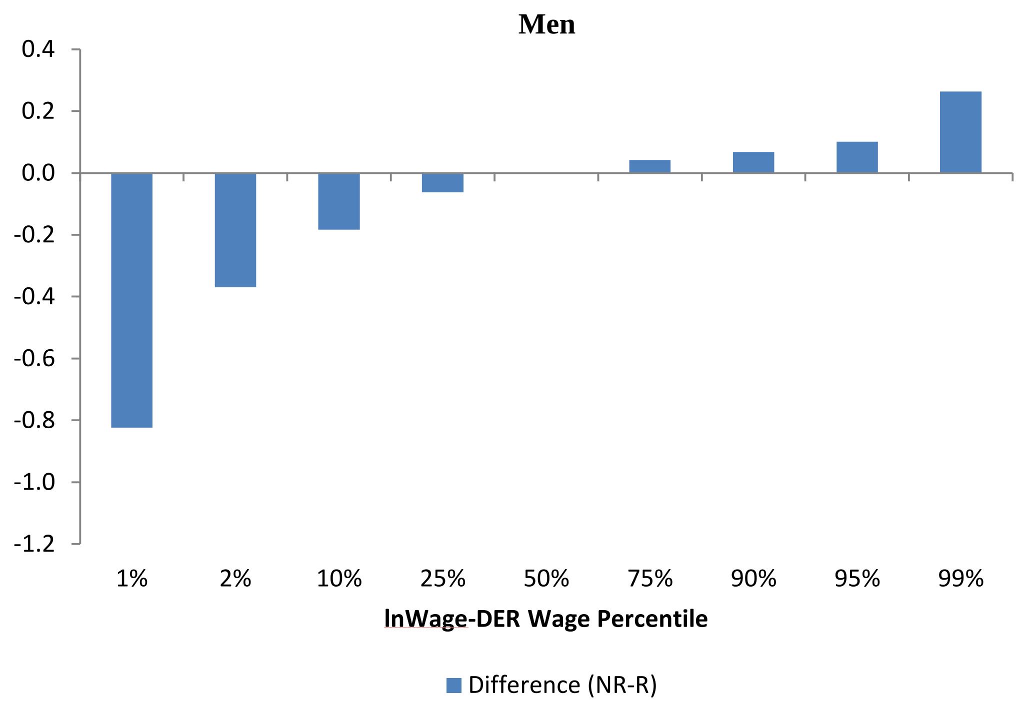
The pattern of response bias is readily seen in Figure 2, which shows differences in DER wage residuals between CPS non-respondents and respondents (NR-R) across the distribution. Evident for men and women is that NR-R differences shift from negative to positive. In lower portions of the distribution we see positive selection into response, with CPS non-respondents having lower DER earnings residuals than respondents. In the middle of the distribution, differences between non-respondents and respondents are effectively zero, indicating little response bias. At the top of the distribution, CPS non-respondents have higher DER wage residuals than do respondents, indicating negative selection into response.[10]

Although our emphasis is on how response bias varies across the distribution, a measure of net bias over the distribution is also of interest. Based on our full-sample log wage regression for men, the mean DER wage residual for CPS non-respondents is -0.011 and that for CPS respondents is 0.019, a -0.031 difference (by construction, the mean residual for the full sample is zero). This indicates that on average there is weak positive selection into response, with male CPS non-respondents having modestly lower DER earnings than respondents, conditional on covariates. Among women, the pattern of positive selection is somewhat stronger. The mean residual for female CPS non-respondents is -0.063 and that for CPS respondents is 0.022, a -0.085 difference as compared to the -0.031 for men.

These net differences in observed DER earnings for CPS respondents and observationally equivalent non-respondents are small, but non-trivial. Based on the 19.5% weighted non-response rate in our male sample, the overall upward bias in mean male CPS earnings due to positive selection would be about 0.6 percent (.195 times -0.031 equals -0.006). For women, bias is a substantive 1.6 percent (.193 times -0.085 equals -0.016). Taken together, this would imply that overall average earnings (for full year/full time workers) are understated by roughly 1 percent due to response bias. Estimates of gender wage gaps are likely to be understated by about 1 percentage point.[11]

[10] For both respondents and non-respondents, wage residuals are mechanically negative (positive) in the left (right) tails of the distribution. Our conclusions are based on *differences* in residuals for respondents and non-respondents.

[11] The downward bias in average earnings is .546 (.006) + .454 (.016) = 0.011, where .546 and .454 are our sample proportions for men and women. Bias in the gender gap is calculated as the difference between 0.006 and 0.016.



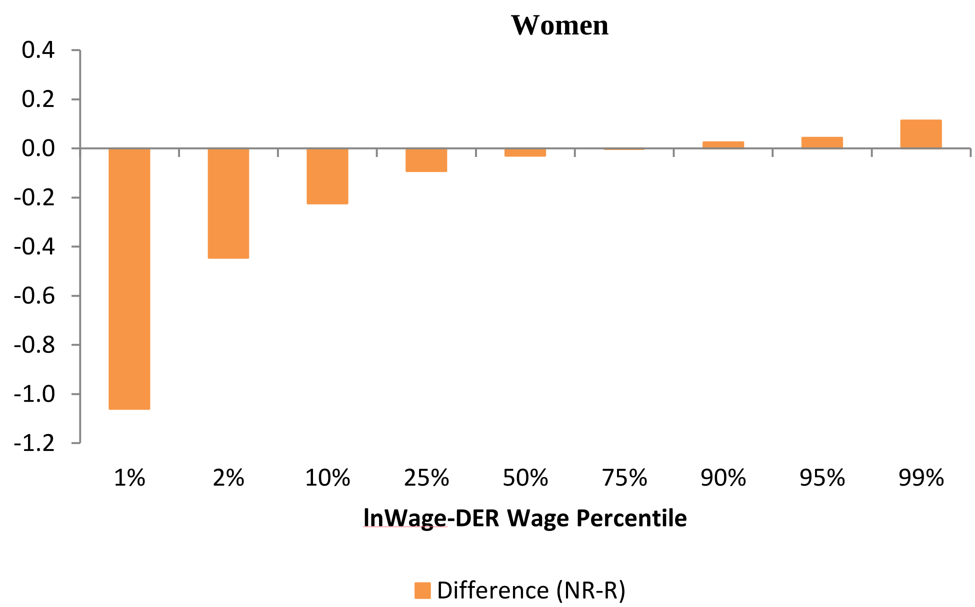


Figure 2: Differences in DER Wage Residuals between CPS Non-respondents and Respondents (NR – R) Across the Distribution, by Sex

See the text for details. For sources, see the note in Table 1. `

7. Additional Evidence and Robustness Checks

In this section, we examine (a) how our results are affected by the sample exclusion of students and those who do not work full-time/full-year; (b) the identification of occupations and worker groups with relatively large shares of earnings off-the books (i.e., not recorded in DER); and (c) the reliability of proxy earnings reports in the CPS.

Sample exclusions. Excluded from our sample were students and those who did not work full time/full year. As a robustness check, we examined whether the non-response pattern for these excluded workers is similar to that seen for our primary sample. We measured non-response rates for these excluded workers, by gender, at each percentile of their DER wage distribution (not shown). Their patterns of non-response are noisy, but both men and women display similar patterns of non-response to those seen for our main samples, with non-response flat over much of the distribution but with evidence of higher non-response in the lower and upper tails. In contrast to results from our primary samples, one does not see extremely high rates of non-response in the lower tail or at the highest percentiles among students and workers who are not FT/FY.

Occupations with off-the-books earnings. We examine the occupations of workers who either are highly likely to have earnings off-the-books or cannot be matched to tax records. Among the occupations with the lowest DER matches are the construction trades (e.g., painters, drywall installers, roofers, brick masons, laborers, and helpers); dishwashers, cooks, dining attendants and bartender helpers, and food preparation workers; grounds maintenance workers; and agricultural and fishing related workers. Using our matched CPS/DER sample, we examine which occupations show the largest percentage (log) gap between CPS and DER earnings. These are typically occupations where workers have some portion of their earnings reported and some off-the-books. Not surprisingly, there is considerable overlap between these occupations and those with the lowest DER match rates. In addition to the types of occupations listed above, we see large CPS-DER earnings gaps for occupations such as real estate brokers and agents, door-to-door sales workers, personal appearance workers, massage therapists, musicians, and bartenders, and clergy.[12] “High-gap” occupations can be categorized as those jobs or types of work where there is often an opportunity to avoid reporting earnings (Roemer 2002).

How serious are off-the-book earnings for our analysis? The short answer is that it is less of a problem than expected. Our concern was that a sizable portion of the non-response seen in the left tail of the DER wage distribution, conditional on earnings attributes, was the result of workers with earnings off the books being reluctant to respond to earnings questions. Our robustness checks were reassuring. When we remove from our male and female samples all workers in the “high gap” occupations (those with large CPS minus DER wage differences) and all foreign-born noncitizens, there is almost total overlap in non-response rates in the left tail (and elsewhere) between these samples and the samples that includes these workers, as shown previously in Figure 1.[13]

[12] Clergy are typically taxed as self-employed workers, but may report earnings in the CPS as wage and salary earnings, thus creating a gap between CPS and DER earnings. Clergy may be exempt from paying taxes on allowances for housing and transportation. They also receive payments for weddings and funerals that may go unreported.

[13] These figures are available on request. Although foreign born noncitizens are disproportionately employed in occupations with high levels of off-the-books earnings, their rates of earnings non-response are lower than among native men and women.

Proxy versus self reports. Roughly half of all earnings reports in the CPS are provided by proxy respondents. And earnings non-response is substantially higher among individuals with a proxy respondent (Bollinger and Hirsch 2013). If one includes proxy dummies in a standard CPS wage equation, one finds substantive negative coefficients associated with the use of non-spouse proxies and coefficients close to zero for spousal proxies. In analysis not reported here, we have used the matched CPS/DER data to examine the quality of proxy earnings reports. The analysis indicates that both spouse and non-spouse proxy reports are accurate, the exception being modest underreporting of married men's earnings by wife proxies (for related evidence, see Reynolds and Wenger 2012). Proxy wage effects found in standard wage equations do not reflect misreporting, but instead worker heterogeneity not captured by standard covariates.

8. Dealing with Non-response: Guidance for CPS Users

The analysis in this paper has implications for researchers using the CPS and similar household data sets such as the American Community Survey (ACS). As discussed earlier, even if non-response is completely missing at random, severe “match bias” can arise in the estimation of earnings equation coefficients if researchers include those with imputed earnings. Attenuation bias is severe for coefficients on variables not used as a hot deck match criterion. Bias is more complex when earnings have been allocated using an imperfect match of donor characteristics. Among the “remedies” for match bias (Bollinger and Hirsch 2006), the simplest and most widely used is to simply throw out imputed earnings and rely on the respondent sample. This sample can be reweighted by the inverse probability of response, but in practice this rarely makes much difference.

The matched CPS-DER data allows us to examine directly whether relying solely on respondents’ earnings produces results similar to those using complete data. Because the DER sample includes administrative earnings for CPS non-respondents as well as respondents, we can compare earnings function parameter estimates from respondent-only samples with those from complete samples, something not possible using the CPS.

Using the DER sample, we have estimated log wage equations with a dense set of covariates, separately for the respondent, non-respondent, and pooled samples. Using estimates from these regressions, we provide the predicted wage for men and women using means from the full CPS sample multiplied by coefficient estimates from these regressions. We use as our benchmark the predicted earnings based on coefficients from the full sample, unobtainable from the CPS because of the absence of non-respondents’ earnings. We compare these full-sample predicted wages to those obtained using the coefficients from the respondent sample, the latter possible to measure with public data.

Focusing first on men, use of full sample coefficients with the full sample worker attributes results in a predicted mean log wage of 2.984. This is close to that obtained using respondent-only coefficients, which leads to a predicted mean log wage of 2.991, or 0.007 (roughly one percent) higher than obtained with the full sample. The equivalent values for women are 2.724 using full sample coefficients and 2.739 using respondent coefficients, a 0.015 difference. These differences reflect a mean tendency toward positive selection into response, more so for women than men. Such selection is more readily evident directly comparing predicted earnings using respondent (R) and non-respondent (NR)coefficients. The R–NR predicted earnings difference is 2.991–2.962 = 0.029 for men and 2.739–2.658 = 0.081 for women. These differences are substantive. Because the non-respondent shares of the total samples are relatively small (roughly 20 percent), the respondent only sample provides coefficient estimates reasonably close to what would be produced using the full sample, the latter not being an option with public use data.

Although our assessment regarding the reliability of respondent-only samples is a positive one, this assessment is based on the accuracy of mean outcomes. As seen in our paper, the news is less rosy in the tails. Bias from non-response prevents researchers from observing many low earners over a fairly wide range and many high earners at the very top of the distribution. The former may be the more serious problem, at least for researchers using public use data. High non-response in the lower tail affects our ability to measure and understand low wage labor markets, low income households, and poverty. Problems in the right tail are concentrated among the very top percentiles, who already have earnings masked (top-coded) in public use files. Research on very high earners is severely constrained, even absent non-response. That said, public use files no doubt include too few top-coded earners due to response bias.

9. Conclusion

This paper addresses the fundamental question of how non-response varies across the earnings distribution, a difficult question to answer and one not adequately examined in prior literature. Using matched household and administrative earnings data, we find that non-response across the earnings distribution, conditional on covariates, is U-shaped, with left-tail “strugglers” and right-tail “stars” least likely to report earnings. Women have particularly high non-response in the left tail; men have high non-response in the far right tail. Using a joint distribution of wages, we see little difference between women and men in non-response at the same wage level. Selection is not fixed across the distribution. In the left tail there is positive selection into response; in the far right tail there is negative selection. Over most of the earnings distribution response bias is ignorable.

Acknowledgments

We thank James Spletzer, Trudi Renwick, Chuck Nelson, Ed Welniak, and Adam Bee for helpful comments, plus participants at presentations at the U.S. Census Bureau, Society of Labor Economists, and university presentations. The analysis in this paper was conducted while Hokayem was employed by the Census Bureau. The views expressed in this research, including those related to statistical, methodological, technical, or operational issues, are solely those of the authors and do not necessarily reflect the official positions or policies of the Census Bureau. The authors accept responsibility for all errors.

References

Abowd, J.M. and M.H. Stinson. 2013. “Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data.” *Review of Economics and Statistics* 95 (5): 1451-1467.

Blackburn, M.L. 2007. “Estimating Wage Differentials without Logarithms.” *Labour Economics* 14 (1): 73-98.

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

Bollinger, C.R. and B.T. Hirsch. 2013. “Is Earnings Nonresponse Ignorable?” *Review of Economics and Statistics* 95 (2): 407-416.

Bond, B.J., D. Brown, A. Luque, and A. O'Hara. 2013. "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey," Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference.

Greenlees, J., W. Reece, and K. Zieschang. 1982. "Imputation of Missing Values when the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association* 77 (378): 251-261.

Herriot, R. A. and E. F. Spiers. 1975. "Measuring the Impact on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources," *Proceedings*, *American Statistical Association*, 147-158.

Hirsch, B.T. and E.J. Schumacher. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation," *Journal of Labor Economics* 22 (3): 689-722.

Larrimore, J., R.V. Burkhauser, S. Feng, and L. Zayatz. 2008. "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976-2007)." *Journal of Economic and Social Measurement* 33: 89-128.

Lillard, L., J.P. Smith, and F. Welch. 1986. "What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation," *Journal of Political Economy* 94 (3): 489-506.

Little, R.J.A. and D.B. Rubin. 2002. *Statistical Analysis with Missing Data, Second Edition*. Wiley-Interscience: Hoboken, NJ.

Nicholas, J. and M. Wiseman. 2009. "Elderly Poverty and Supplemental Security Income," *Social Security Bulletin* 69 (1): 45-73.

Reynolds, J. and J.B. Wenger. 2012. "He Said, She Said: The Gender Wage Gap According to Self and Proxy Reports in the Current Population Survey," *Social Science Research* 41 (2): 392-411.

Roemer, M. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in the Current Population Survey and the Survey of Income and Program Participation." LEHD Program Technical Paper No. TP-2002-22, US Census Bureau.

Rubin, D.B. 1976. "Inference and Missing Data (with Discussion)," *Biometrika* 63: 581-592.

Suits, D.B. 1984. "Dummy Variables: Mechanics V. Interpretation," *The Review of Economics and Statistics* 66 (1): 177-180.

Welniak, E.J. 1990. "Effects of the March Current Population Survey's New Processing System On Estimates of Income and Poverty," *Proceedings, American Statistical Association*.