The Good, The Bad and The Ugly: Measurement Error, Non-response and Administrative Mismatch in the CPS.*

Christopher R. Bollinger[†], Barry T. Hirsch[‡], Charles M. Hokayem[§], James P. Ziliak[¶]

August 2018

Abstract

Using the March Current Population Survey Annual Social and Economic Survey matched to the Social Security Administration Detailed Earnings Records, we link observations across year to investigate a relationship between item non-response and measurement error for the CPS earnings questions. Prior research has found that (1) non-response is linked to earnings: individuals in the tails of the earnings distribution are less likely to respond to the earnings question. Other research has suggested that (2) individuals with income above the average are more likely to under-report their earnings, while individuals with earnings below average are more likely to over-report their earnings. We examine whether these two phenomenon are related. The overlapping samples in the CPS data allow us to observe individuals who switch from response to non-response. This allows us to investigate whether those who fail to respond in both years have different response patterns than those who provide earnings data in both years.

*This work was conducted under Special Sworn Status (Bollinger, Hirsch, Ziliak) at the University of Kentucky Research Data Center. The views expressed here are the opinions of the authors and do not reflect those of the U.S. Census Bureau. This work is preliminary and incomple.

†Corresponding Author: University of Kentucky, Department of Economics, Gatton College of Business, Lexington, KY 40503; crboll@uky.edu.

‡Georgia State University

§U.S. Census Bureau

¶University of Kentucky

1 Introduction

Survey data is crucial for social science research. Only survey data can achieve both a broad collection of variables and population representativeness. For example, research into the determinants of earnings requires measures of earnings, but also measures of education, labor market experience, gender, and race. Creative new analyses require the addition and presence of new measures beyond those typically included in the basic Mincerian wage specification. Moreover, survey data are available over long periods of time. For example, the Current Population Survey has been collected in some form since March of 1964. Administrative data such as tax data may contain measures of earnings or income, but does not contain even the most basic demographic variables. However, survey data suffers from data quality issues such as measurement error and item non-response. A variety of literature has focused upon measurement error in survey reports of earnings (Mellow and Sider, 1983; Duncan and Hill, 1985; Bound and Krueger, 1991; Bound et al., 1994; Pischke, 1995; Bollinger, 1998; Bound, Brown and Mathiowetz, 2001; Roemer, 2002; Kapteyn and Ypma, 2007; Meijer et al., 2012; Abowd and Stinson, 2013). Although there are exceptions, most studies find support for the "common man" hypothesis: that low income individuals tend to over-report earnings, while high income individuals tend to under-report earnings. Kapteyn and Ypma (2007), Meijer et al. (2012) and Abowd and Stinson (2013) call into question the typical assumption that the administrative record is perfectly measured. These studies find support that administrative data may have match error or measurement error. We also find evidence that challenges the "Administrative Gold Standard" assumption. This in turn suggests that the "common man" hypothesis is less evident, or even nonexistent.

A growing literature considers item non-response in survey data. As noted initially in Hirsch and Schumacher (2004), the rate of item non-response to the earnings questions in the ASEC (and the Monthly Outgoing Rotation question), rose dramatically through the 90's and especially the early 2000's. In the 1980's, the non-response rate hovered around 12 to 15%. During the 1990's the rate rose and through the mid 2000's and 2010's was as high as 20% or more. (See Bollinger and Hirsch, 2006). Item non-response means that while a survey participant generally answers other questions in both the monthly and the ASEC supplement, that individual refuses to respond to certain questions. One of the highest rates of item non-response in the CPS are the questions about labor market earnings. There are many possible reasons for the refusal. One possible reason is simply ignorance. The interview structure of the CPS means that nearly 50% of all responses are proxy responses. The CPS asks the household to designate a single individual as the respondent, rather than separately interviewing each member of the household. Earnings non-response for Proxy responses is - on average - substantially higher than for respondents. Other reasons for non-response are stigma or threat. Stigma may occur if individuals feel embarrassment about their earnings: either because they are too high, or they are too low, relative to some perceived norm. Threat can occur for a variety of reasons such as tax evasion or simply fear of release of sensitive information.

Initial work on earnings non-response in the CPS (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006) established that the Census imputation procedure led to bias in regression coefficients unless researchers limited their specification to include only the variables used in the procedure, measured in the same way as used in the procedure. These findings led to the recommendation that researchers drop imputed earnings from such measures. Alternatively, researchers may use a selection model to address non-response, or construct an imputation approach consistent with the underlying research model of interest (see Little and Rubin, 2014). One can also rebalance the sample using inverse probability weights. Bollinger and Hirsch (2006) find that in practice, these approaches have little impact on the results. Bollinger and Hirsch, (2013), Bollinger et al. (2018), and Valet et al (2018) investigated the implications of the resulting sample selection and found that non-response was concentrated on those individuals with low earnings or high earnings. Others have examined entire survey non-response, including Bee, at al. (2015) and Meyer and Mittag (2015). Like the "Common Man" finding, non-response to the earnings question is concentrated among those with either low earnings or high earnings: non-response is in the tails of the distribution. This concentration of measurement error and non-response among the same groups of earners, suggests there may be a relationship between non-response and measurement error.

A number of authors have considered the possibility of a relationship between survey non-response and measurement. Bollinger and David (2001) find a relationship between response error in Food Stamps in the first waves of the 1984 SIPP and subsequent attrition from the sample. They hypothesize a "good reporter - bad reporter" type phenomenon: individuals who engage with the survey provide accurate responses and remain in the sample. Those who do not engage have responses that contain errors and are likely to fail to respond to the survey at all. Similar hypotheses have been forwarded as far back as Cannel and Fowler (1963) and Cochoran et al (1954). Dominitz and Manski (2017) examine the potential trade-off between improved response rates and measurement error. A number of authors (Groves 2006; Olsen 2007; Abraham, Helms and Presser, 2009) examine how correlation between a variable of interest and response propensity would effect non-response bias. Another direction of this research classifies respondents as "reluctant" when it takes survey enumerators multiple calls and discussion to obtain an interview. Examples of this include, Kreuter, Müller and Trappmann (2010), Triplett et al. (1996), Stoop (2005), Dahlhammer, Simile and Taylor (2006) and Fricker (2007). Nicoletti, Perecchi, and Foliano (2011) establish bounds for poverty rates allowing for very general missing and non-response patterns. The work here differs from this previous work in that we have individuals who both respond and don't respond to the earnings question in two different time periods. In many ways this provides a cleaner definition of "reluctant responder" than previous work, and focuses on the specific question. The strength of previous work is often measuring overall survey willingness compared to measurement within the survey.

We investigate and find some support for a similar finding here. We make use of the March Current Population Survey - Annual Social and Economic Survey (CPS-ASEC) matched to W-2 records of earnings. This allows us to observe earning for individuals who do not report earnings in the CPS, as well as those who do. The CPS sample structure allows a two year panel of individuals. We focus on the sample of those who the CPS follows for two years. This allows us to observer multiple opportunities to respond to the earnings question. Perhaps unsurprisingly, the vast majority of respondents report their earnings in both years. However, nearly 20% of respondents switch from response to non-response or vice versa, and nearly symmetrically so. Those who otherwise participate in the survey but fail to report their earnings in both years are the smallest of the four possible groups. Thus for nearly 20% of the sample, we can observe response in one period, and non-response in the other. It is comparing this group to those who respond in both periods that allows us to address the question of whether non-response and reporting error may be linked. Although at times we treat the administrative record as the "Gold Standard", when that assumption is relaxed, the structure of the measurement error is less clear. However, the relationship between the measurement error and the non-response is supported through a variety of models and estimation strategies.

The main finding is important in a number of respects. First, it suggests that attempts to cajole or otherwise improve response from non-responders maybe less valuable than previously thought. If these individuals are failing to provide quality data, it may be best to simply allow them the freedom to refuse. Secondly, it suggests that using their responses to proxy for their other missing data may not be wise. Further, it suggests that assumptions of "random" response error and random non-response are problematic. The concentration of both non-response and response err in the tails of the distribution suggest that perhaps these individuals have higher costs (psychic or simply recall) in providing data.

2 Data

The data derive from the 2006 through the 2011 March Current Population Annual Social and Economic Survey (CPS-ASEC). The ASEC is important in that it is the source of the official U.S. poverty rate, one of the most common data sets to use for examining income distributions and inequality, the workhorse for understanding the determinants of earnings and the impact of policies on earnings, and heavily used for research on participation in public safety net and other similar programs. The monthly Current Population Survey is administered to approximately 60,000 households every month. The monthly survey is designed to measure labor market activity - in particular unemployment - for the U.S. The survey is structured so that the address is the sampling unit. The address is chosen and contacted, and remains in the sample, initially, for four consecutive months. Thus an address chosen and initially contacted in January, would appear in the January, February, March and April monthly survey. One year later the same address is contacted again, and included in the sample for those same months in subsequent years. Earnings information is collected in two ways, for households in the fourth month and eighth month (the so called "outgoing rotation groups") of their survey time period. This measure is focused on hourly and weekly earnings rather than annual earnings. In March, the Annual Social and Economic Survey is administered. Among a wide variety of additional questions, earnings from all employment, and details on the industry and occupation of the primary employer, are elicited.

Using internal files, the CPS-ASEC data are matched to earnings data from the Social Security Administration (SSA). We utilize the Detailed Earnings Records (DER) file provided by the SSA. The file has a "Personal Identification Key" for each individual which uniquely identifies that individual. Although not the individual's Social Security Number (SSN), it plays a similar role and is based upon the master identification file kept by the SSA. The U.S. Census uses name, address, birth date, gender, and SSN to identify individuals and obtain their "PIK." We were granted access to a file which allows us to match the PIK to individuals in the CPS, and then to use that to match individuals to the SSA DER. Our final match rate is 89.6% for men and 91.4% for women. As discussed in Bollinger et al. (2018), non-matched individuals are more likely to be foreign born, and are more likely to have lower educational attainment. They also report lower earnings in the CPS.

We use 2006 through 2011 data (reflecting incomes in 2005 through 2010). Prior to 2006, the Census Bureau followed an "opt in" strategy for linking survey to administrative data: individuals had to agree, by responding that they would allow a Census to link their responses to administrative data. After 2006, Census adopted an "opt out" strategy where individuals had to refuse being linked to administrative data. While the burden on the respondents was not particularly high, linkage rates rose by over 5% between 2005 and 2007. We use the higher link rate data for this analysis. At this writing, DER records were not available after 2011.

In addition we focus on full time, full year workers between the ages of 18 to 65. The concentration of non-response in the tails of the earnings distribution is most pronounced for this group (although hourly measures for the full sample were equally pronounced). Most earnings determination research focuses on full year, full time workers in this age range and we follow this convention. Because of the rotation structure of the CPS, any individual appearing in the ASEC, can potentially appear in the ASEC the next year. Using standard CPS household and individual identifiers, we construct the sample of individuals who are linked across response years. The CPS is not an individual or family based sample, but rather an address based sample. Thus individuals who move from their original address are not followed by the CPS the next year, and hence cannot be linked. The linked ASEC sample is somewhat selective: it tends to be older, more highly educated, have a higher concentration of whites and married individuals than the full sample (see Bollinger et al. 2018). It has also been found (Bollinger, 1998) that this group has lower measurement error (as measured by variance). We separate men and women throughout, but do not address sample selection into full year, full time. Thus the final analysis sample consists of full year, full time workers between ages 18 and 65, who have been linked across two years of the census, and who have been matched to the SSA DER records. We note that this sample is not representative of the U.S. population, or even the U.S. full time, full year workers. Hence we do not use sampling weights in this initial analysis. This sample allows us to investigate measurement error and non-response through comparison to the DER earnings over multiple years.

Table 1 provides mean and standard deviation of common demographic variables (included in the models estimated below) and earnings in the first CPS year contact (denoted throughout "year 1"). We note here that for men, CPS earnings are, on average, lower than DER earnings, while for women the reverse is true, but the difference is much smaller. Caution should be used, since non-respondents (who have imputed earnings in the CPS earnings) are included in the averages. Due to disclosure restrictions, sample sizes are rounded to the nearest 500 observations. This sample is slightly older, more white, more married and more highly educated than the population. We also note this group is more likely to be Native born. Overall, 52.8% of men and 40% of women are "proxy" respondents. Proxy respondents are concentrated among spouses (wife providing information for the husband and others in the household or vice versa). As discussed in Bollinger and Hirsch (2013), spousal proxies are more likely to respond to the earnings question than non-spousal proxies, but less likely than self reports.

Table 2 provides details on response. Overall, in the sample, the non-response rate for men in the first year is 17.0%, while the non-response rate for women is 15.9% Second year non-response rates are 17.9% and 16.8%. While non-response does rise between the first and second year, the more interesting aspect is that individuals switch status. Only 7.2% of men and 6.5% of women fail to respond in both years: less than half of non-respondents in year one are non-respondents in year two as well. Fully 10.8% of men and 10.3% of women respond in year one, but then refuse to respond in year two. Perhaps surprisingly, 9.8% of men and 9.4% of women refuse to respond in year one but then respond in year two. If non-response were concentrated among those who fail to respond in both years, investigation of links between non-response and measurement error would be tenuous, or in the extreme, impossible. It is the approximately 20% (20.6% of men and 19.8% of women) of the sample who switch between non-response and response that will drive our analysis.

Table 3 presents log earnings and log earnings differentials across the four response types. While Census provides imputations, we do not report these in this table to highlight that we do not have earnings reports for these individuals. Initially, let us consider the second column of each gender panel, where individuals responded in both years. We note low average real earnings growth reported in the CPS of about 1% for women and an actual decline of about 1% for men. The data span the great recession; hence, the low growth is expected. Interestingly, for both men and women, the DER growth is higher: for women it is 3% growth, while for men it is 0% growth. The response difference is the log of the CPS earnings report minus the log of the DER earnings report. We note that the difference in earnings reports for men in year 1 is fully 5%, but falls to 4% in year two. For women, it is 4% in year 1 and falls to 2% in year two; both men and women over-report their earnings on average. When we examine the two columns which represent those who switch response, we note that differences between CPS and DER earnings are higher (particularly for men) in each year. Men who report in year one report 6% higher earnings than are recorded in the DER (on average) and those report only in year two report 7% higher earnings than recorded in the DER. Similarly for women, the differences are 4% and 3%. We also note that for both men and women, those who respond in year two have the highest DER earnings growth (5% for women and 3% for men) while those who report only in year 1 have the lowest earnings growth (2% for women and -3% for men). This finding is similar to that of Bollinger et al. (2018).

Table 4 presents four simple regressions: CPS earnings on DER earnings for those who respond in the year given, and a Probit model of CPS earnings response on DER earnings and the difference between CPS earnings and DER earnings measured in the other year for those who responded in the other year. In the CPS earnings model, only year indicators were included for other controls and none were statistically significant. Most measurement error literature has found little correlation between reported earnings and typical demographic variables once administrative earnings (DER) are controlled for (for a review, see Bound, Brown and Mathiowetz, 2001). Estimates including other variables using these data were found to be qualitatively similar to those reported here. In the non-response model, however, the presence of other variables have strong - and well known - explanatory power and are included to ensure that we are isolating the relationships with earnings. Estimates without control variables are also similar.

In the first and third column of each gender panel we note the usual "common man" finding: the coefficient on the administrative earnings measure (the DER earnings) is less than one. We also report the variance of the error term (the mean squared error) as an estimate of the measurement error variance. In the case of no measurement error, the coefficient on earnings would be one, and the variance of the error term would be zero. As these two summary measures differ from the ideal of 1 and 0, one can argue that "more error" is apparent. We note that women have less error than men: the coefficients on DER earnings are close to one and the Mean Squared error is smaller. Other researchers have noted this finding (Bound and Krueger, 1991; Bollinger, 1998; Bound at al., 2001).

In the second and fourth column for each gender in table four we present the results from estimation of a Probit model on response (1 = responded in the period). It is important to understand the sample used here. In models of response for year one, in order to have a measure of the measurement error we use individuals who responded in year 2. Response for year 1 is measured for those who responded to both years or to the second year only. This provides a measure of the difference between a survey response and the administrative record. In addition to variables reported here, we include controls for labor market experience, education, race, census region, metropolitan size and year. We first note that in all cases the earnings response is clearly related to the DER earnings reports for that year: response in year one (two) is related to DER Earnings measure in year one (two). The negative on the squared earnings component supports the results of Bollinger et al (2018) which finds that non-response is highest in the tails. We use the difference between the CPS-reported Earnings and the DER Earnings to measure the measurement error. We first note that the level of measurement error is insignificant across all specifications. However, except for the women in year one, the squared difference is significant and negative. Thus the larger (in magnitude) the difference between the CPS report and the DER report, the less likely the individual will have responded in the opposite year. We report these relationships simply as a starting point for evidence that measurement error predicts non-response.

3 Models of Response Error and Non-Response

In this section we propose two models which relate measurement error and non-response. The first model is a conceptual model, meant to fix ideas on why the two processes may be linked. The second is a statistical model which provides the basis for the analysis below.

3.1 Conceptual Model of Response Effort

Groves (2006), Groves et al (2004), Beimer et al (2004), and Groves et al (2002) provide a broad overview of both non-response mechanisms and measurement error mechanisms. The simple behavioral model we present here attempts to provide an understanding of how effort in response may result in both the common man type response error and non-response. We also provide some evidence for higher nonresponse among those with higher measurement error variance when they do respond. Our departure is a simple equation for "cost" in responding to a question (or potentially a survey):

$$Cost = k + s\left(Y-\mu\right)^2 + c\left(Y-Y^*\right)^2.$$

The respondent chooses Y, the response; while Y* is the true value of the variable. The term k is the baseline "cost" of recalling/responding. The term $s\left(Y^*-\mu\right)^2$ is the "stigma" of being either a high or low earner relative to the mean μ of the population. Hence, small s would imply that the respondent does not feel much stigma about earnings relative to the population mean. The term $c\left(Y-Y^*\right)^2$ is the respondent's mental cost of providing poor answers. A large c would indicate someone who desires to be helpful to the survey, while a small c would indicate a respondent who is unengaged and uninterested in response accuracy. Throughout, we assume that c and s are both non-negative and that $(c+s)>0$. The survey respondent will seek to minimize the cost of participation in choosing either to respond at all, and the Y they will respond with. The first order conditions of the problem of choosing Y to minimize response cost yield:

$$2s\left(Y-\mu\right) + 2c\left(Y-Y^*\right) = 0.$$

Solving this for Y produces:

$$Y = \frac{s\mu}{s+c} + \left(\frac{c}{s+c}\right)Y^*,$$

which establishes a linear relationship between the response Y and the true variable, Y^*. We note that the coefficient $0 \leq \left(\frac{c}{c+s}\right) \leq 1$ for any finite choice of c and s. The coefficient will equal 1 if $s=0$, which implies then that $Y=Y^*$. If the respondent feels no stigma about either being above or below "average" they will choose to correctly respond. The coefficient will equal 0 if $c=0$, in which case the respondent will simply provide "the mean" as a response. This results in the "common man" type response relationship. We further assume that if total cost is too high, the individual will fail to respond completely:

$$NR = 1\left[k + s\left(Y - \mu\right)^2 + c\left(Y - Y^*\right)^2 > 0\right].$$

Substitution of the optimal response into the formula yields:

$$NR = 1\left[k + s\left(\frac{s\mu}{s+c} + \left(\frac{c}{s+c}\right)Y^* - \mu\right)^2 + k\left(\frac{s\mu}{s+c} + \left(\frac{c}{s+c}\right)Y^* - Y^*\right)^2 > 0\right].$$

Gathering terms and simplifying produces a simple u-shaped relationship between non-response and true earnings, as found in Bollinger at al. (2018):

$$NR = 1\left[k + \frac{sc}{s+c}\left(Y^* - \mu\right)^2 > 0\right].$$

Note here, then, if $s = 0$ or $c = 0$, nonresponse is determined only by k. We posit that any finding of a relationship between non-response and measurement error implies (c, s) both positive. We also note that the model can be extended to be "statistical" in nature by allowing (k, s, c) to be random variables.

3.2 Statistical Model of Data Generation

Building on the simple relationships above, we posit a model of the data generating process consistent with that of Kapteyn and Ypma (2007) and Abowd and Stinson (2013) which allows a variety of possible special cases that we consider and discuss in the results section. We begin by assuming that log-earnings are determined by a standard Mincerian type wage equation:

$$Y_{it}^* = X_{it}\beta + u_{it}, \tag{1}$$

where Y_{it}^* is person i's log-earnings in time period t. Here, t will refer to either the first or second year in the March Survey. The X_{it} are standard explanatory variables including potential experience, education, race,gender etc. The term u_{it} is a term meant to capture unobserved factors which determine earnings. The ideal Y_{it}^* is not directly observed. Rather we observe two different measures of that:

$$Y_{it}^D = \begin{cases} Y_{it}^* + \varepsilon_{1it}^D & \text{with prob. } p \\ \mu_Y + \varepsilon_{2it}^D & \text{with prob. } 1-p \end{cases} \tag{2}$$

$$Y_{it}^{C}=\begin{cases}\delta_1+\rho_1 Y_{it}^{*}+\varepsilon_{1it}^{C} & \text{with prob. } q\\ \delta_2+\rho_2 Y_{it}^{*}+\varepsilon_{2it}^{C} & \text{with prob. } 1-q\end{cases} \tag{3}$$

The first measure, Y_{it}^{D} is the earnings from the DER (the administrative records). The second measure, Y_{it}^{C} is the survey report from the CPS. The two models for Y_{it}^{D} represent both a mismeasured version (the first equation) and a mismatched version (the second equation). The data for X_{it} derive solely from the match to the CPS. Hence, if a mismatch occurs we expect no correlation between the observed (survey) X_{it} and the observed Y_{it}^{D}. Hence in the second equation we model the data as a random draw from the entire distribution of earnings. The model for the CPS measure of earnings, Y_{it}^{C} posits two measurement error models allowing for different types of response. The two models are very general, and allow for a variety of response types. As noted above, the "severity" of the measurement error problem is often summarized in the two parameters $(\rho,\sigma_{\varepsilon}^{2})$. Hence we posit that $|\rho_1-1|>|\rho_2-1|$ and $\sigma_{\varepsilon 1}^{1}>\sigma_{\varepsilon 2}^{2}$: those in group 2 are "better reporters" than those in group 1. The empirical question we seek to answer is whether we find "good reporters" and "bad reporters." We hypothesize that non-response can help us discern who is in which group a-priori.

We also posit a response model for the survey data.

$$R_{it}=\begin{cases}1 \quad if \quad Z_{it}\gamma+h\left(Y_{it}^{C},Y_{it}^{D},Y_{it}^{*}\right)+v_{it}>0\\ 0 \qquad\qquad otherwise\end{cases}. \tag{4}$$

The variable R_{it} is a 1 if person i had a response to the earnings questions in the March CPS in year t. Here, $t=1$ or $t=2$ represents the first or second appearance in the CPS March survey. The variables Z_{it} are known covariates with non-response including proxy response status,education, race, gender and age. The function $h\left(Y_{it}^{C},Y_{it}^{D},Y_{it}^{*}\right)$ measures a variety of potential relationships between non-response and the various measures of income. We leave this very general at this point, but note that missing at random assumes that $h\left(Y_{it}^{C},Y_{it}^{D},Y_{it}^{*}\right)=0$, while prior work (Bollinger et al, 2018) found a U-shaped function in Y_{it}^{D} . That is if "good reporters" have less measurement error (as posited above) they should also be the ones more likely to respond to the survey.

Different assumptions on this model have led to different estimates of both the relationships between Y_{it}^D and Y_{it}^C as well as non-response. Much of the classical measurement error literature (see for example, Bound and Krueger, 1992; Bollinger, 1996; Bound et al, 2001) assumes that $p = 1, q = 1$, and $V\left(\varepsilon_{1it}^D\right) = 0$: that is that the administrative records are equivalent to the true earnings (Y_{it}^*) and that there is one simple summary model of misreporting. Bollinger (1996) showed that a simple linear model may not be appropriate, however, Kapteyn and Ypma (2007) and Abowd and Stinson (2013) suggest that if bad matches are allowed $(p < 1)$ the linear model fits well. Kapteyn and Ypma (2007) provide some evidence that the bad matches seem to explain much of the "common man" hypothesis: on average, low earners tend to over report, while high earners tend to under-report (implying $\rho_1 < 1$).

Bollinger and David (2001), while not considering income, suggest that there is a relationship between measurement error in a variable and non-response. They also suggest that there may be two types of responders, those who are engaged with the survey, attempt to give accurate responses and continue to participate over time and across questions and those who not engaged and are the primary source of both non-response and measurement error. Our model above attempts to capture all of these factors, but as such is complicated, and difficult to estimate. We note that in fact, the structure of the non-response equation - and the findings of Bollinger and David (2001) would indicate further that q may be a function of R_{it}: measurement quality is related to non-response.

Our approach here is to consider a variety of restrictions on the models for estimation. The comparison of the models provides some insight into the underlying measurement process. Rather than claiming we have found the right model (as is often done) we explore and examine a variety of models.

An important consideration in interpreting the parameters of the measurement error and non-response models is bias in use of public use CPS data. Most researchers only observe Y_{it}^{C}. Understanding the measurement process allows improved modeling. When the CPS earnings are used as a dependent variable, the coefficient ρ impacts the bias of slope coefficient estimates: $\rho = 1$ implies no bias. The variance of ε_{it}^{C} only impacts the standard errors and the fit of the model. Indeed, there is a balance: removing the "highest" ε_{it}^{C} may not reduce standard errors because it necessarily means reducing the sample size. Bollinger and Chandra (2005) also point out that the common approaches of trimming and windsorizing may indeed induce bias even if it does not exist. Bollinger and Hirsch (2013) examine whether there is selection bias due to non-response. Bollinger and Hirsch (2006) establish that using census imputations for non-response leads to bias in all but a very narrow set of cases. Again, balance here is paramount: the small bias to selection appears to be the "lesser" of two evils. Further, we may be able to model the selection process more easily.

In cases where the CPS earnings are used as an explanatory variable in a regression, both ρ and $V\left(\varepsilon_{it}^{C}\right)$ impact the bias. It is possible that certain combinations lead to no bias, but this is generally serendipitous. Here though, typically, the classical measurement error effect $V\left(\varepsilon_{it}^{C}\right)$ dominates. Papers by Bollinger (2003), Hyslop and Imbens (2001), Ditraglia and Garcia-Jimeno (2015) and suggest some possible approaches, including bounding and instrumental variables.

4 Estimation

Identification of the parameters in the above models highly depends upon the assumptions made. A variety of authors have sought to estimate models on measurement error, and different assumptions have led to different estimates. This paper seeks to investigate whether there is a relationship between reporting error among CPS earnings respondents and non-response to the CPS earnings question. Hence, we estimate the measurement error model (and subsequently non-response models) under a variety of assumptions seeking to establish that relationship across a variety of models. Our results do not appear to settle the question of whether the administrative record is measured without substantial error, or may have serious mismatch issues (as has been done by Katpeyn and Ypma, 2007, among others), but rather to understand the relationship between response errors and non-response. Nonetheless, we find some evidence for mismatch and measurement error in the DER (administrative) measures of income.

4.1 Model 1: DER Records assumed correct

We begin by estimating the standard type models often seen in the measurement error literature: the administrative record is taken as "correct" while the survey data are taken to be mismeasured. Formally, we are assuming that $Y_{it}^{D} = Y_{it}^{*}$, and thus in equation 2, $p = 1$ and $V\left(\varepsilon_{1it}^{D}\right) = 0$. This is similar to the results in Table 4 where we further assumed only one type of measure error structure ($q = 1$ in equation 3).These assumptions imply that the regression of Y_{it}^{C} on Y_{it}^{D} identifies ρ as the slope coefficient and $V\left(\varepsilon_{it}^{C}\right)$ as the variance of the residuals. We begin by relaxing the assumption of only one type of CPS measurement error structure (see equation 3): first we consider NR_{it} to perfectly determine which group (using non-response in one year to indicate "poor reporter" in the other year), then we consider a simple finite mixture model, and finally a finite mixture model where R_{it} is used as a covariate in the mixture probability.

Table 5 presents the measurement error model (equation 3) estimates under the assumption that the DER measure is correct. We estimate the model conditional on response status. The first two columns in each gender panel present the results of the OLS regression of the CPS earnings measure on the DER earnings measure (and indicators for CPS years, which are not reported, and generally insignificant) for the first and second response year, for those individuals who responded in both years ($R_{it} = 1$ for all $t = 1, 2$). The third and fourth columns represent those who only responded in year 1 or year 2 respectively. These can be compared to the columns in table 4, where all individuals are pooled together. The vast majority of individuals respond in both years, and hence those coefficients and variances are very similar to the results in Table 4. However, for those men who respond only in year 1, the estimate for ρ is lower (further from 1) at 0.634 in year 1 and 0.602 in year 2 (as compared to 0.699 and 0.700 for those who respond in both years). Standard hypothesis tests comparing the two coefficients reveal test statistics of 5.48 and 8.27, thus the difference is certainly statistically significant. Additionally, the variance of the error for those men who only respond in one of the two years is 0.394 and 0.349 as compared to 0.208 and 0.166 for those who respond in both years. F-tests comparing the two variance estimates yield test statistics of 1.90 and 2.11, again rejecting the null of equal variances.

For women the story is virtually identical. As noted above, women's estimates of ρ are uniformly closer to one and their estimates of the error variance are smaller than the men's. However, we again find the pattern where ρ is further from one and the error variances are larger for those who only respond only in one of the two years. Women who respond in both years have estimates for ρ of .764 and .760, while women who only respond in one year have estimates of 0.655 and 0.681. Tests for the difference in these two estimates have test statistics of 7.36 and 6.63, again rejecting the null of similarity. Tests for differences in the variances have F-test statistics of 2.53 and 1.36, again rejecting the null of same variance. As hypothesized, this provides some evidence that individuals who respond to the survey have less measurement error than those who do not.

Although the above results suggest that non-response does have some predictive power for the quality of data, it is both quite likely and quite possible that there are "good reporters" who for other reasons do not respond to the survey in both year, or "bad reporters" who do respond in both years. One approach to separating individuals into these two categories is to estimate a finite mixture model (FMM). Finite mixture models hypothesize exactly the type model presented in the CPS model described in the previous section. Estimation of finite mixture models is based on a maximum likelihood estimator, where the probability of an individual being from either category is estimated:

$$L=\sum_{i=1}^{n}\left(f_1\left(Y_{it}^{C}|Y_{it}^{D}\right)*q+f_2\left(Y_{it}^{C}|Y_{it}^{D}\right)(1-q)\right).$$

Here, we assume that $f_,$ is a normal distribution. There are two approaches taken here to estimate the models: in the first, we only allow q to be a single parameter. This is parsimonious and allows the data to dictate the two underlying distributions. The drawback is that FMM will "find" two groups. What these two groups represent is up to interpretation of the researcher. While it may be reasonable to infer that they represent two response styles, one might postulate other explanations. The other approach estimates a logit model for q. We use the response status as the only explanatory variable in this case. While again, interpretation of the two groups is left to the researcher, the importance of the non-response in determining the probability of being a member of the two groups provides some rigor in interpretation.

Table 6 presents the simple model where q is estimated as a single parameter. The first component we provisionally label as the "good reporters:" The coefficient on the DER earnings in this component is closer to one and the variances are smaller than component two. The second component has much lower estimates for ρ and much higher estimates for the variance of the error and so represent the "bad responder." Indeed in all cases, the estimates are more extreme than those found in table five. While suggesting two different response types, this approach doesn't establish any relationship between the two groups and the responder category.

In table 7 we re-estimate the model, but allow the probability of being in each group to depend upon the response status. That is when estimating the components for year one, the probability q is a function of response status (responder or non-responder) in year two. The estimation model uses a logit type specification for q and we report the coefficient on the response status in the column for component 1. However, it should be noted it is not predicting the CPS earnings, but rather which component an individual is most likely to be associated with. Initially note the nearly identical results we find with this approach compared to those in table 6 where the classification is a single parameter: the coefficients in each component are nearly identical, the variances are nearly identical. In table 7 we report the average of the predicted probabilities as the "predicted proportion" while in table 6, the estimate of the single parameter q (and 1-q) is reported. Even these are nearly identical. The estimated coefficients on response are positive, large and highly statistically significant: response predicts which component an individual's response is classified into. The marginal effects vary between 0.128 and 0.186.

In table 8 we present models on non-response as a function of measures of the measurement error models estimated above (see equation 4). Like those models in table 4, response for year one is estimated on the sample of respondents in year 2 and vice versa. We also include controls for experience, education, race, census geography, and metropolitan size. As in table 4, the inverted U-shaped relationship between response and the DER earnings in that year remains. In the first two columns we use the residual from the regressions in table 4. These are comparable to the residuals from table 5, except they do not separate by response (the dependent variable). It should be noted that in many ways, this misses the systematic part of measurement error measured by ρ. In both years, for men, we find - as in table 4 - that the level of the residual is not statistically significant, but the square of the residual is negative and statistically significant: the higher the magnitude of the error, the less likely someone will be a respondent in the opposite year. Like the results in table four, the results for women using the residual are somewhat mixed. The signs follow the same patterns as for the men, but the coefficient on residual squared is not significant for women's response in the first year. It is, however, statistically significant in the second year.

Additionally we use measures generated by the Finite Mixture Model results. We use the results from table 6, which does not use the response status to estimate the model. We argue that "component one" is the lower measurement error category and we use two measures. First we use the posterior predicted probability that an individual would be classified as a member of component one. In both years for both men and women, this variable is highly significant and positive: likelihood of being a member of the component one group is associated with higher probability of being a respondent. Similarly, we use the predicted indicator. The indicator is one if the individual's probability of being in component one is greater than their probability of being in component two (essentially if their predicted probability is larger than 0.5). Again, the coefficient is statistically significant and positive. We again interpret this as better reporting in the other period leads to higher likelihood of response.

We conclude here that there is evidence for two models of earnings response and that response status (non-response) does have power to predict the group to which an individual belongs. We also find evidence that the size of the error variance has power to predict response status and that the prediction of the measurement group to which an individual belongs has power to predict response status. We take this as evidence in favor of a "good reporter/bad reporter" type phenomenon.

4.2 Model 2: DER records with additive white noise error

Several authors have provided evidence that administrative earnings may not be without error. A number of sources of error can occur, off the books earnings being the most common. This would imply that $p = 0$ still, but allows $V\left(\varepsilon_{it}^{D}\right) > 0$ in equation 2. In this case, the regression of Y_{it}^{C} on Y_{it}^{D} (as in the previous section) would result in estimates of ρ that are biased toward zero: the classical measurement error bias result. Assuming the model here, though, a simple IV estimator is easily motivated. The regression of Y_{it}^{D} on X_{it} will produce consistent estimates of the parameters β in the Mincerian model. The additive error term does not impact the consistency of those parameters. Hence, we can rewrite equation 3 (the model for

Y_{it}^{C}) as

$$Y_{it}^{C} = \delta + \rho Y_{it}^{*} + \varepsilon_{1it}^{C} = \delta + \rho\left(X_{it}\beta + u_{it}\right) + \varepsilon_{it}^{C} \qquad (5)$$
$$= \delta + \rho\left(X_{it}\beta\right) + \left(\rho u_{it} + \varepsilon_{it}^{C}\right) = \delta + \rho\widehat{Y}_{it}^{D} + \left(\rho u_{it} + \varepsilon_{it}^{C}\right).$$

The error term $\left(\rho u_{it} + \varepsilon_{it}^{C}\right)$ is uncorrelated with $\widehat{Y}_{it}^{D}$, the predicted value from the regression of Y_{it}^{D} on X_{it}. Note, however, that this estimator - like all IV estimators - is consistent even if there is no measurement error in Y_{it}^{D}. Hence, if there is no measurement error in the DER data, the estimated coefficients from the regression of Y_{it}^{C} on Y_{it}^{D} should not differ from the estimated coefficients from the regression of Y_{it}^{C} on $\widehat{Y}_{it}^{D}$. Appendix Table 1 provides the estimates from the first stage regression, while table 9 provides estimates of the simple model (similar to that provided in table 4). In all four models (two years, two genders), the coefficients rise to over 0.9. This is consistent with findings of Katpeyn and Ypma (2007) and suggests that the common man hypothesis may not be supported, or not supported strongly. We reject the hypothesis that the estimated ρ's in table 9 are equal to one in all cases.

To estimate the measurement error variance in equation 3, we note that the residuals from the first stage regression of Y_{it}^{D} on X_{it} are

$$e_{it}^{D} = u_{it} + \varepsilon_{it}^{C}.$$

Similarly, the residual from the regression of Y_{it}^{C} on $\widehat{Y}_{it}^{D}$ are given in equation 5 as,

$$e_{it}^{C} = \rho u_{it} + \varepsilon_{it}^{C}.$$

Thus

$$V\left(e_{it}^{C}\right) = V\left(\rho u_{it} + \varepsilon_{it}^{C}\right) = \rho^{2} V\left(u_{it}\right) + V\left(\varepsilon_{it}^{C}\right)$$

and

$$Cov\left(e_{it}^{C}, e_{it}^{D}\right) = \rho V\left(u_{it}\right).$$

We are assuming that $\varepsilon_{it}^{C}, \varepsilon_{it}^{D}$ are uncorrelated with each other and with u_{it}. We note that the data generating process for each measure are different, and so argue that the assumption of uncorrelation seems plausible. Similar assumptions have been made by Kaptyn and Ympa (2007) and Abowd and Stinson (2011????). Thus

$$V\left(\varepsilon_{it}^{C}\right) = V\left(e_{it}^{C}\right) - \rho Cov(e_{it}^{C}, e_{it}^{D}).$$

In tables in this section we provide estimates of both $V\left(e_{it}^{C}\right)$ (mean squared error) and $V\left(\varepsilon_{it}^{C}\right)$ (error variance)[1]. Note that in table 9, the mean squared error is larger than in table 4, as we would expect using the predicted earnings. While the estimated variance term in table 9 is smaller than the mean squared error in table 4 (which, under the assumptions in the previous section should estimate the error variance). F-tests were all at least 1.6, and so reject the hypothesis of equality at conventional levels. We take this as some evidence that measurement error in the DER earnings is present.

In table 10, we re-estimate the measurement error models of table 5, this time using the predicted earnings from the regression of DER earnings on typical earnings regression variables (see Appendix table 1 for details). Again, we use non-response in the other year to separate the individuals who respond in both years from those who respond in only one year. Here we find a remarkable switch in coefficients. In table 10 we find that those who only respond in one year have an estimated ρ closer to one than those who respond in both years, except for the women in year two. However, except for the men in year two, the difference between the estimates for respondents in both years are not statistically different from the estimates for respondents in one year at conventional levels. The test statistic for men in year two is -2.11, rejecting the null of equality, however, the other tests are -1.65, -1.04, and 0.59. We note that the mean squared error is uniformly higher for those who only respond in one year. Our estimates for the error variance term follow a similar pattern. Our estimated variances are nearly twice as high for the switchers as they are for those who respond in both years. Only the women in year two are closer. In all four comparisons, the

[1] There is a second approach, informed by the IV literature. The adjusted residuals, $e_{it}^{CA} = Y_{it}^{C} - \rho Y_{it}^{D} = \varepsilon_{it}^{C} - \rho \varepsilon_{it}^{D}$. Then, $V\left(\varepsilon_{it}^{C}\right) = V\left(e_{it}^{CA}\right) - \rho^2 V\left(e_{it}^{D}\right) + \rho Cov(e_{it}^{D}, e_{it}^{C})$. These estimates are nearly identical to those provided with the formula above.

F test statistics comparing the variances range from 1.32 to 3.12, and thus are well above the critical values of 1.05[2]. We take this as evidence of a relationship between response and measurement error, but here the focus is on the variance.

In Table 11 we re-estimate the FMM using the predicted earnings from the first stage DER regression. We use the response status again in estimating the component probabilities. We do not report estimates of the model where the component probabilities are a simple constant, because the results are nearly identical (as they were in tables 6 and 7). Component 1 of all four models has the highest probability in the sample. In all four models, the estimate of ρ in component one is less than one while the estimate of ρ in component two is larger than one. The component one estimates of ρ are further from one - lower - than the estimates in table 10. In three of the four models, the estimate of ρ for component one is closer to one than in component two, often quite substantially. The exception is year two for the men where component two has an estimate of ρ of 1.067 compared to 0.912 for component one. We hypothesize that component one still represents the "good reporters" as the coefficients on the response status in the probability models remain positive and statistically significant. The marginal effects estimates, however, are markedly lower, ranging from 2% for the two female estimates to 4% for the two male estimates. The coefficients themselves are somewhat smaller. We note too that overall the probabilities of being in the "good reporter" category are much higher. One interpretation of these results is that some of the observed "poor reporting" is due to errors from the DER, rather than errors from the CPS. We note that component 1, in all four cases, has both a much lower mean squared error and a much lower estimate of the measurement error variance than component two. The FMM splits the sample on these variances quite strongly. We also note that while evidence for the "common man" hypothesis is still present, it is much weaker. Moreover, the worst reporters seem to follow a much different approach with understatement of low earnings and

[2] Similarly, f-tests of the mean squared error range from 1.23 to 1.74, well above the critical values.

overstatement of high earnings. We should caution that mixture models simply split the data into two models which average to the common model through the simple probability model also estimated. The interpretation of the meaning of those models is left to the researcher and the reader.

In table 12 we provide estimates of response as a function of the measurement error models in this section. As in section 4.1, we use three measures: the residuals from the regressions in table 9 where the predicted log DER earnings is used, the predicted probability of being in component 1 and the predicted indicator of component 1 in the FMM models. Here we simply use the residuals, e_{it}^{C} as described above. We have also used the adjusted residuals ($Y_{it}^{C} - \rho Y_{it}^{D} = \varepsilon_{it}^{C} - \rho \varepsilon_{it}^{D}$) with nearly identical results. The FMM model used to construct the predicted probability and the predicted indicator for component 1 does not include response in the component probabilities model (as in the results reported in table 11). Rather we use a model like that in table 6, but with the predicted DER earnings, as used elsewhere in this section. For the men in table 12, again in both cases, we find that the squared residuals are negative and highly significant. For response in year two, the level of the residual was also statistically significant at the two-tailed 10% level. The coefficient itself is negative. In both cases, the larger the magnitude, the lower the response probability. We also note those whose measurement error is negative (those who under-report their earnings) have higher response probability (holding constant the magnitude) than those whose residual is positive. Like the results in table 8, the coefficients on the measures from the FMM estimates show that higher likelihood of being in component one (or an indicator for that), yields higher likelihood of responding. For women in year one, we find that the residuals are not statistically significant in predicting response. In year two, the usual pattern emerges with squared residuals being negative and statistically significant at the 5% two tailed level. The level is not significant, as was generally true in table 8. As with the residual results, the probability of being in component one in the FMM is only statistically significant for the year two response model, although both have the positive sign. Moreover, the indicator variable is not significant for either year.

We find less evidence for the common man hypothesis in this section, but we continue to find evidence of a link - in particular for men - between non-response and measurement error. Men and women who respond to the survey have lower measurement error variance than those who do not, while certainly men with higher measurement error variance have lower response rates to the earnings question. While the results are somewhat weaker for women, this may be due to the composite construction of the measurement error. In future research, we intend to further extend this section to more carefully isolate the measurement error component.

4.3 Model 3: Allowing for both measurement error and mismatch in the DER earnings

The final set of models we estimate allow for mismatch in DER earnings as well as measurement error. In some respects, the approach in section 4.2 allows for that as well, essentially allowing for heteroskedasticity associated with an unobserved "match" variable. In this section we use Finite Mixture Models to estimate the earnings regression with DER earnings in the first stage. This allows q equation 2 to be estimated, as well as β in equation 1. The earnings are predicted from the "matched" component. We initially estimated an unrestricted model, where slope coefficients were estimated for each component. We found that the estimated coefficients were similar for the two models. It should be noted that this was the finding in the unrestricted model. In our approach here, we force the coefficients in the second component to be zero, following the assumptions implied by equation 2 and administrative mismatch hypothesis. This imposes the "mismatch" interpretation on the data. It is not clear that an unrestricted model supports that data generating process.

Appendix table 2 presents the estimated coefficients from the DER wage model.

Many of the coefficients - in particular those associated with the education and race variables - are similar to estimates from the simple OLS estimates in appendix table one. We do note that some coefficients - in particular those on the potential experience variables - are markedly different.

Table 13 is comparable to tables 4 and 9. As in table 9, we use the predicted value of earnings from the DER regression. In this section we use the prediction from the FMM model where DER is the dependent variable. The results are very similar to table 9: estimates of ρ range from 0.92 to 0.96, much higher than in table 4. With the exception of the coefficient for men in the first year, all other coefficients are within 0.006 of the counterparts in table 9. Even the coefficient for men in the first year differs only by 0.02. We interpret this to imply that the FMM model for the first stage does not substantially change the estimates.

Table 14 is comparable to table 10 (and table 5): we estimate the measurement error models by response status using the predicted wages from the first stage FMM model. Like the relationship between tables 13 and 9, there is little difference in the estimates of the parameter ρ in table 14 as compared to table 10 where the simple OLS regression is used for the prediction of earnings. The largest difference is between the coefficients for men who respond only in the first year and that differences is 0.02. The other estimates are between 0.002 and 0.006 in difference. While these vary in statistical significance, the economic significance is low for all. Again, it appears that there is little gain in estimating the FMM for the first stage: the simple IV estimate performs well.

Finally in table 15 we again estimate a FMM for the measurement error model, using the prediction from the FMM first stage. The results are comparable to those found in table 11. Again, we find little difference using the first stage FMM model over using a simple linear regression estimated with OLS. The conclusion concerning the relationship between measurement error and non-response remains similar: non-response does appear to have predictive power for which of the two measurement models an individual belongs, with non-response associated with the model having the highest variation and generally - although less clear than in table seven - an estimated ρ further from one. What we conclude in this section is that a model that attempts to allow for mismatch in the DER earnings regressions does not improve upon simply using the predicted DER from the first stage regression.

Table 16 examines the models for response using the approach in this section. The predicted value for DER earnings derives from the first component of the DER earnings models, the measurement error residuals derive from the models estimated in table 13. The two measures from the FMM models derive from a model where the probability of being in component one (lower measurement error) is not a function of the response. As we have seen in both of the subsections above, the squared residual for men is negative and highly statistically significant in both years, and the level of the residual is not significant. Both measures of component probability are significant. The more likely an individual is to be in component one, the more likely they are to respond in the opposite year. For women though, the results are weak, as they were in section 4.2. As in the previous section, only the squared residuals in year two were statistically significant and again, only at the 10% level. Similarly, the FMM probabilities are only significant for year two, and again only at the 10% level.

5 Conclusions

The principle result of our analysis is that there is evidence that non-response and measurement error are related: individuals who fail to respond to the earnings questions in the survey in one year, have higher measurement error than those who respond in both years; those who appear to have higher measurement error are less likely to respond to the survey. This has a number of implications both for researchers and for survey design. It is not clear, though, whether working to improve either response or reduce measurement error in the survey is cost effective. This paper does not investigate this, but does suggest that any approach which takes only response rates or only measurement error into account, is missing an important link.

Researchers who seek to improve the data quality might select on the sample of only those who respond in both years. Prior research by Bollinger and Hirsch (2006,2013) establishes that using Census imputations generally leads to bias in a regression context particularly for coefficients on variables not matched in the Census imputations. However, research by Bollinger and Hirsch (2013) and Bollinger et al. (2018) suggest that in some circumstances the sample selection associated with non-response will lead to bias in estimates. The results here however do seem to indicate that - at least for the case where earnings is used as a dependent variable - the bias on coefficients of including the mismeasured data is likely to be low. Bias is determined only by ρ and the best estimates - those in section 3.2 - suggest that it is close to 1 on average, indicating attenuation bias of 10% or less. Indeed, the evidence from both sections 3.2 and 3.3 suggest that the bias in the setting when the earnings are used as a dependent variable may be lowest for those who respond only in one year. The cost will be higher residual variance and thus larger standard errors and lower R-squared.

References

[1] Abraham, Katharine G., Sara Helms, and Stanley Presser (2009). “How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States.” American Journal of Sociology 114:1129–65.

[2] Abowd, John M. and Martha H. Stinson (2013). "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," Review of Economics and Statistics, Vol XCV, no. 5, pp. 1451-1467.

[3] Beimer, Paul P., Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz and Seymour Sudman (2004). Measurement Error in Surveys. John Wiley and Sons, Inc. New Jersey.

[4] Bee, Adam, Graton M. R. Gathright and Bruce D. Meyer (2015) "Bias from Unit Non-Response in the Measurement of Income in Household Surveys" Working Paper

[5] Bollinger, Christopher R. (1998) "Measurement Error in the CPS: A non-parametric Look," Journal of Labor Economics, Vol. 16, no. 3, pp 576-594.

[6] Bollinger, Christopher R. and Martin H. David (2001) "Measurement Error and Nonresponse in the SIPP" Journal of Business and Economic Statistics, Vol 19., pp 129-141.

[7] Bollinger, Christopher R. and Amitabh Chandra (2005) "Iatrogenic Specification Error: Cleaning Data can Exacerbate Measurement Error Bias" Journal of Labor Economics, Vol. 23, no. 2. pp. 235-257.

[8] Bollinger, Christopher R. and Barry T. Hirsch (2006) "Match Bias in the Earnings Imputations in the Current Population Survey: The Case of Imperfect Matching" Journal of Labor Economics, vol 24. no. 3, pp 483-520.

[9] Bollinger, Christopher R. and Barry T. Hirsch (2013) "Is Earnings Response Ignorable?" Review of Economics and Statistics, Vol 95, no 2. pp. 407-416.

[10] Bollinger, Christopher R., Barry T. Hirsch, Charles Hokayem, and James P. Ziliak (2018) "What have we learned?" Journal of Political Economy, forthcoming.

[11] Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rogers (1994). "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," Journal of Labor Economics, Vol 12. pp. 345-368.

[12] Bound, John, Charles Brown, and Nancy Mathiowetz (2001). "Measurement Error in Survey Data" in J.J. Heckman and E. Leamer, eds., Handbook of Econometrics (New York: Elsevier), pp. 3705-3843.

[13] Bound, John and Alan B. Kreuger (1991) "The Extent of Measurment Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics, Vol 9, pp. 1-24.

[14] Cannell, Charles F., and Floyd J. Fowler (1963). “Comparison of a Self-enumerative Procedure and a Personal Interview: A Validity Study.” Public Opinion Quarterly 27:250–6

[15] Cochran, W., F. Mosteller and J. Tuckey (1954). Statistical Problems of the Kinsey Report on Sexual Behavior in the Human Male (Washington, DC: American Statistical Association).

[16] Dahlhamer, James M., M. Catherine Simile, and Beth Taylor (2006). “Exploring the Impact of Participant Reluctance on Data Quality in the National Health Interview Survey (NHIS).” In Proceedings of the Statistics Canada Symposium 2006: Statistical Issues in Measuring Population Health. Catalogue No. 11-522-XIE. Gatineau, Quebec, Canada.

[17] DiTraglia, Francis J. and Camilo Garcia-Jimeno (2015) "A Framework for Eliciting, Incorporating, and Discipling Identification Beliefs in Linear Models), working paper.

[18] Duncan, Greg J. and Daniel H. Hill (1985). "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data," Journal of Labor Economics, Vol 3, pp. 508-532.

[19] Fricker, Scott S. (2007). "The Relationship between Response Propensity and Data Quality in the Current Population Survey and the American Time-use Survey." PhD diss., University of Maryland.

[20] Groves, Robert M. (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys." Public Opinion Quarterly 70:646–75.

[21] Groves, Rober M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau (2004). Survey Methodology. John Wiley & Sons, Inc: New Jersey.

[22] Groves, Rober M., Don A. Dillman, John L. Eltinge, and Roderick J.A. Little (2002). Survey Nonresponse. John Wiley & Sons, Inc: New Jersey.

[23] Hirsch, Barry T. and Edward J. Shumacher (2004) "Match Bias in Wage Gap Estimates Due to Earnings Imputation" Journal of Labor Economics, Vol 22, no. 3, pp. 689-722.

[24] Hyslop, Dean R. and Guido W. Imbens (2001) "Bias From Classical and Other Forms of Mesurment Error," Journal of Business and Economic Statistics, vol 19, no.4, pp 475-481.

[25] Katpeyn, Arie and Jelmer Y. Ypma (2007). "Measurement Error and Misclassification: A Comparison of Survey adn Administrative Data," Journal of Labor Economics, Vol 25, pp. 513-550.

[26] Kreuter, Frauke, Gerrit Müller, and Mark Trappmann (2010). "Nonresponse and Measurement Error in Employment Research" Public Opinion Quarterly, Vol. 74, no. 5, pp 880-906.

[27] Little, Roderick J. and Donald B Rubin (2014) Statistical Analysis with Missing Data, John Wiley & Sons: New Jersey.

[28] Manski, Charles F. and Jeff Dominitz (2017) "More Data or Better Data? A Statistical Decision Problem," Review of Economic Studies, vol 84. pp 1583-1605.

[29] Meijer, Erik, Susann Rohwedder, and Tom Wansbeek (2012). "Measurement Error in Earnings Data: Using a Mixture Model Approach to Combine Survey and Register Data," Journal of Business and Economic Statistics, Vol 30, pp. 191-201.

[30] Meyer, Bruce D. and Nichola Mittag (2015) "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net," Pe

[31] Mellow, Wesley and Hal Sider (1983). "Accuracy of Response in Labor Market Surveys: Evidence and Implications," Journal of Labor Economics, Vol 1, pp.331-344.

[32] Nicoletti, Cheti, Franco Peracchi and Francesca Foliano (2011)"Estimating Income Poverty in the Presence of Missing Data and Measurement Error" Journal of Business and Economic Statistics, Vol.29, no. 1, 61-72.

[33] Olsen, Kristin (2007). "An Investigation of the Nonresponse–Measurement Error Nexus." PhD diss., University of Michigan.

[34] Pischke, Jörn-Steffen (1995) "Measurement Error and Earnings Dynamics: Some Estimates from teh PSID Validation Study," Journal of Business and Economic Statistics, Vol 13, pp. 305-314.

[35] Roemer, Marc (2002) "Using Administrative Earnings Records to Assess Wage Date Quality in the March Current Population Survey and the Survey of Income and Program Participation," LEHD Technical paper TP-2002-22.

[36] Stoop, Ineke A. L (2005). The Hunt for the Last Respondent. The Hague: Social and Cultural Planning Office (SCP).

[37] Triplett, Timothy, Johnny Blair, Teresa Hamilton, and Yun Chiao Kang (1996). "Initial Cooperators vs. Converted Refusers: Are There Response Behavior Differences?" Paper presented at the Annual Conference of the American Association for Public Opinion Research, Salt Lake City: Utah.

[38] Valet, Peter, Jule Adriaans and Stefan Leibig (2018) "Comparing Survey data and Administrative records on gross earnings: nonreporting, misreporting, interviewer presence adn earnings inequality" Quality and Quantity, on line: https://doi.org/10.1007/s11135-018-0764-z

Table 1: Sample Means for Matched and Linked Sample

	Men		Women	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
Age	43.11	10.98	43.05	10.92
White	0.768	0.422	0.742	0.437
Black	0.0680	0.252	0.0971	0.296
Asian	0.0410	0.198	0.0424	0.201
Other	0.0248	0.156	0.0258	0.159
Hispanic	0.0982	0.298	0.0924	0.290
Education Years	13.94	2.689	14.10	2.489
Married Spouse Present	0.716	0.451	0.644	0.479
Married Spouse Absent	0.109	0.312	0.188	0.391
Never Married	0.175	0.380	0.168	0.374
Native Born	0.886	0.317	0.896	0.305
Foreign Born- Citizen	0.0592	0.236	0.0603	0.238
Foreign Born Non-Citizen	0.0545	0.227	0.0436	0.204
Proxy Respondent	0.528	0.499	0.402	0.490
Spouse Proxy	0.397	0.489	0.284	0.451
NonSpouse Proxy	0.131	0.337	0.118	0.322
DER Earnings	63840	75860	39400	43160
CPS Earnings	66080	166900	39030	58890
Sample Size (rounded)	41000		30000	

Table 2: Earnings Item Response Rates for Matched Linked Sample

	Male	Female
Non-Respond Both Years	7.2	6.5
Respond Both Years	72.2	73.8
Respond in Year 1 Only	10.8	10.3
Respond in Year 2 Only	9.8	9.4
First year non-response	17.0	15.9
Second year non-response	17.9	16.8
Switchers	20.6	19.8

Table 3: Log Earnings Means and Differences

	Men				Women			
	Non-Respond Both Years	Respond Both Years	Respond Year 1	Respond Year 2	Non-Respond Both Years	Respond Both Years	Respond Year 1	Respond Year 2
Log CPS Earnings Year 1		10.74	10.66			10.25	10.19	
Log CPS Earnings Year 2		10.73		10.65		10.26		10.21
Log DER Earnings Year 1	10.63	10.69	10.60	10.55	10.16	10.21	10.15	10.12
Log DER Earnings Year 2	10.64	10.69	10.57	10.58	10.18	10.24	10.17	10.17
Log Response Difference Year 1		0.05	0.06			0.04	0.04	
Log Response Difference Year 2		0.04		0.07		0.02		0.03
Log CPS Earnings Growth		-0.01				0.01		
Log DER Earnings Growth	0.01	0.00	-0.03	0.03	0.02	0.03	0.02	0.05
Rounded Sample Sizes	3500	30000	4000	3500	2500	22000	3000	2500

Table 4: Simple Measurement Error and Non-Response Model Estimates

	Men				Women			
	Log CPS Earnings Year 1	Earnings Response Year 1	Log CPS Earnings Year 2	Earnings Response Year 2	Log CPS Earnings Year 1	Earnings Response Year 1	Log CPS Earnings Year 2	Earnings Response Year 2
Log DER Earnings Year 1	0.689***	0.807***			0.747***	1.053***		
	(0.00349)	(0.107)			(0.00383)	(0.194)		
Log DER Earnings Year 1 squared		-0.0356***				-0.0469***		
		(0.00506)				(0.00941)		
Log DER Earnings Year 2			0.686***	0.741***			0.749***	0.828***
			(0.00316)	(0.110)			(0.00361)	(0.175)
Log DER Earnings Year 2 squared				-0.0319***				-0.0374***
				(0.00519)				(0.00849)
Log Earnings Difference Year 1				-0.00426				0.0194
				(0.0169)				(0.0244)
Log Earnings Difference Year 1 squared				-0.00887***				-0.0181***
				(0.00290)				(0.00501)
Log Earnings Difference Year 2		-0.0150				0.0268		
		(0.0181)				(0.0273)		
Log Earnings Difference Year 2 squared		-0.0181***				-0.00343		
		(0.00374)				(0.00632)		
Mean Squared Error	0.231		0.186		0.154		0.132	
Education, race, region, metro size	No	Yes	No	Yes	No	Yes	No	Yes
Year Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size (Rounded)	34000	33500	33500	34000	24500	30000	30000	24500

*** p<0.01, ** p<0.05, * p<0.1, models of Earnings include year indicators, models of non-response include proxy indicators, experience, education, race, metro size, census division and year indicators.

Table 5: Simple Measurement Error Models by Response Status

	Men				Women			
	Respond Both Years		Switchers		Respond Both Years		Switchers	
	Log CPS Earnings Year 1	Log CPS Earnings Year2	Log CPS Earnings Year 1 Only	Log CPS Earnings Year 2 Only	Log CPS Earnings Year 1	Log CPS Earnings Year2	Log CPS Earnings Year 1 Only	Log CPS Earnings Year 2 Only
Log DER Earnings Year 1	0.699***		0.634***		0.764***		0.655***	
	(0.00361)		(0.0113)		(0.00382)		(0.0143)	
Log DER Earnings Year 2		0.700***		0.602***		0.760***		0.681***
		(0.00322)		(0.0114)		(0.00380)		(0.0113)
Mean Squared Error	0.208	0.166	0.394	0.349	0.130	0.127	0.329	0.172
Sample Size (Rounded)	30000	30000	4000	3500	22000	22000	3000	2500

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 6: Finite Mixture Models for Measurement Error

	Men				Women			
	Log CPS Earnings Year 1		Log CPS Earnings Year 2		Log CPS Earnings Year 1		Log CPS Earnings Year 2	
	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
Log DER Earnings Year 1	0.949***	0.425***			0.962***	0.454***		
	(0.00163)	(0.00977)			(0.00177)	(0.0123)		
Log DER Earnings Year 2			0.959***	0.434***			0.961***	0.442***
			(0.00150)	(0.00786)			(0.00178)	(0.0115)
Mean Squared Error	0.12	0.885	0.109	0.72	0.112	0.761	0.111	0.681
Estimated Proportion	0.769	0.231	0.732	0.268	0.796	0.204	0.794	0.206
Sample Size (Rounded)	34000		33500		25000		24500	

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 7: Finite Mixture Model for Measurement Error, Mixture Probability on Response Status

	Men				Women			
	Log CPS Earnings Year 1		Log CPS Earnings Year 2		Log CPS Earnings Year 1		Log CPS Earnings Year 2	
	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
Log DER Earnings Year 1	0.950***	0.426***			0.962***	0.454***		
	(0.00163)	(0.00973)			(0.00177)	(0.0123)		
Log DER Earnings Year 2			0.959***	0.434***			0.962***	0.443***
			(0.00150)	(0.00784)			(0.00178)	(0.0115)
Respondent in year 1			0.766***				0.666***	
			(0.0463)				(0.0583)	
Respondent in year 2	0.747***				0.607***			
	(0.0438)				(0.0546)			
Mean Squared Error	0.12	0.884	0.109	0.719	0.112	0.761	0.11	0.679
Predicted Proportion	0.768	0.232	0.787	0.213	0.796	0.204	0.792	0.208
Sample Size (Rounded)	34000		33500		25000		24500	

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 8a: Probit Models of Non-Response on Measurement Error - Men

	Men					
	Response Year 1	Response Year 2	Response Year 1	Response Year 2	Response Year 1	Response Year 2
Log DER Earnings Year 1	0.825***		0.640***		0.672***	
	(0.107)		(0.107)		(0.107)	
Log DER Earnings Year 1 squared	-0.0362***		-0.0282***		-0.0295***	
	(0.00509)		(0.00509)		(0.00509)	
Log DER Earnings Year 2		0.750***		0.523***		0.566***
		(0.110)		(0.108)		(0.108)
Log DER Earnings Year 2 squared		-0.0322***		-0.0227***		-0.0245***
		(0.00522)		(0.00513)		(0.00514)
OLS Measurement Error Residuals Year 1		-0.0167				
		(0.0200)				
OLS Residuals Year 2 Squared		-0.0101***				
		(0.00334)				
OLS Measurement Error Residuals Year 2	-0.0329					
	(0.0214)					
OLS Residuals Year 2 Squared	-0.0202***					
	(0.00443)					
FMM Comp. 1 Predicted Probability Year 1				0.331***		
				(0.0250)		
FMM Comp 1 Predicted Probability Year 2			0.329***			
			(0.0253)			
FMM Comp 1 Predicted Indicator Year 1						0.260***
						(0.0219)
FMM Comp 1 Predicted Indicator Year 2					0.259***	
					(0.0218)	
Sample Size	33500	34000	33500	34000	33500	34000

*** p<0.01, ** p<0.05, * p<0.1, all models include proxy indicators, experience, education, race, census region, city size and year.

Table 8b: Probit Models of Non-Response on Measurement Error - Women

	Women					
	Response Year 1	Response Year 2	Response Year 1	Response Year 2	Response Year 1	Response Year 2
Log DER Earnings Year 1	1.074***		0.803***		0.830***	
	(0.193)		(0.190)		(0.190)	
Log DER EARnings Year 1 squared	-0.0482***		-0.0361***		-0.0373***	
	(0.00944)		(0.00930)		(0.00930)	
Log DER Earnings Year 2		0.865***		0.683***		0.705***
		(0.175)		(0.170)		(0.170)
Log DER EARnings Year 2 squared		-0.0393***		-0.0316***		-0.0324***
		(0.00856)		(0.00828)		(0.00831)
OLS Measurment Error Residuals Year 1		0.0126				
		(0.0276)				
OLS Residuals Year 2 Squared		-0.0171***				
		(0.00544)				
OLS Measurement Error Residuals Year 2	0.0325					
	(0.0305)					
OLS Residuals Year 2 Squared	-0.00126					
	(0.00720)					
FMM Comp. 1 Predicted Probability Year 1				0.245***		
				(0.0316)		
FMM Comp 1 Predicted Probability Year 2			0.253***			
			(0.0332)			
FMM Comp 1 Predicted Indicator Year 1						0.204***
						(0.0288)
FMM Comp 1 Predicted Indicator Year 2					0.200***	
					(0.0277)	
Sample Size	24500	25000	24500	25000	24500	25000

*** p<0.01, ** p<0.05, * p<0.1, all models include experience, education, race, census region, city size and year.

Table 9: Measurement Error Models using Predicted DER Earnings

	Men		Women	
	Log CPS Earnings Year 1	Log CPS Earnings Year 2	Log CPS Earnings Year 1	Log CPS Earnings Year 2
Predicted Log DER Earnings Year 1	0.906***		0.960***	
	(0.00836)		(0.00951)	
Predicted Log DER Earnings Year 2		0.919***		0.963***
		(0.00804)		(0.00909)
Mean Squared Error	0.371	0.321	0.278	0.249
Estimated Error Variance	0.143	0.093	0.096	0.072
Sample Size Rounded	34000	33500	25000	24500

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 10: Measurement Error Models using Predicted DER Earnings by Response Status

	Men				Women			
	Respond Both Years		Switchers		Respond Both Years		Switchers	
	Log CPS Earnings Year 1	Log CPS Earnings Year2	Log CPS Earnings Year 1 Only	Log CPS Earnings Year 2 Only	Log CPS Earnings Year 1	Log CPS Earnings Year2	Log CPS Earnings Year 1 Only	Log CPS Earnings Year 2 Only
Predicted Log DER Earnings Year 1	0.900***		0.948***		0.954***		0.992***	
	(0.00868)		(0.0277)		(0.00970)		(0.0352)	
Predicted Log DER Earnings Year 2		0.911***		0.975***		0.965***		0.946***
		(0.00828)		(0.0292)		(0.00950)		(0.0305)
Mean Squared Error	0.347	0.304	0.540	0.473	0.255	0.243	0.445	0.299
Estimated Error Variance	0.127	0.081	0.258	0.187	0.069	0.061	0.213	0.081
Sample Size (Rounded)	30000	30000	4000	3500	22000	22000	3000	2500

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 11: Finite Mixture Model for Measurement Error Predicted Log Earnings.

	Men				Women			
	Log CPS Earnings Year 1		Log CPS Earnings Year 2		Log CPS Earnings Year 1		Log CPS Earnings Year 2	
	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
Predicted Log DER Earnings Year 1	0.897***	1.227***			0.949***	1.434***		
	(0.00690)	(0.170)			(0.00812)	(0.217)		
Predicted Log DER Earnings Year 2			0.912***	1.067***			0.950***	1.541***
			(0.00712)	(0.128)			(0.00807)	(0.203)
Respondent in year 1			0.674***				0.543***	
			(0.117)				(0.169)	
Respondent in year 2	0.780***				0.579***			
	(0.110)				(0.152)			
Mean Squared Error	0.218	4.951	0.208	3.063	0.18	3.706	0.172	2.866
Estimated V(e-cps)	0.04	2.915	0.029	1.428	0.025	1.612	0.022	0.787
Predicted Proportion	0.969	0.031	0.96	0.04	0.974	0.026	0.973	0.027
Sample Size Rounded	34000		33500		25000		24500	

Table 12a: Probit Models of Non-Response on Measurement Error - Men

	Men					
	Response Year 1	Response Year 2	Response Year 1	Response Year 2	Response Year 1	Response Year 2
Log DER Earnings Year 1	0.799***		0.780***		0.815***	
	(0.110)		(0.109)		(0.108)	
Log DER EARnings Year 1 squared	-0.0339***		-0.0337***		-0.0354***	
	(0.00546)		(0.00523)		(0.00517)	
Log DER Earnings Year 2		0.712***		0.661***		0.693***
		(0.113)		(0.111)		(0.110)
Log DER EARnings Year 2 squared		-0.0294***		-0.0278***		-0.0293***
		(0.00552)		(0.00528)		(0.00525)
IV Measurment Error Residuals Year 1		-0.0353*				
		(0.0213)				
IV Residuals Year 1 Squared		-0.0118***				
		(0.00334)				
IV Measurement Error Residuals Year 2	-0.0343					
	(0.0234)					
IV Residuals Year 2 Squared	-0.0169***					
	(0.00441)					
FMM_IV Comp. 1 Predicted Probability Year 1				0.328***		
				(0.0693)		
FMM IV Comp 1 Predicted Probability Year 2			0.233***			
			(0.0725)			
FMM-IV Comp 1 Predicted Indicator Year 1						0.253***
						(0.0632)
FMM-IV Comp 1 Predicted Indicator Year 2					0.156**	
					(0.0652)	
Sample Size	33500	34000	33500	34000	33500	34000

*** p<0.01, ** p<0.05, * p<0.1, all models include experience, education, race, census region, city size and year.

Table 12b: Probit Models of Non-Response on Measurement Error - Women

	Women					
	Response Year 1	Response Year 2	Response Year 1	Response Year 2	Response Year 1	Response Year 2
Log DER Earnings Year 1	1.075***		0.965***		1.014***	
	(0.198)		(0.198)		(0.195)	
Log DER EARnings Year 1 squared	-0.0494***		-0.0430***		-0.0454***	
	(0.00996)		(0.00966)		(0.00952)	
Log DER Earnings Year 2		0.882***		0.807***		0.834***
		(0.179)		(0.176)		(0.174)
Log DER EARnings Year 2 squared		-0.0404***		-0.0364***		-0.0377***
		(0.00903)		(0.00861)		(0.00851)
IV Measurment Error Residuals Year 1		0.0152				
		(0.0295)				
IV Residuals Year 1 Squared		-0.0111**				
		(0.00505)				
IV Measurement Error Residuals Year 2	0.0367					
	(0.0332)					
IV Residuals Year 2 Squared	-0.000407					
	(0.00694)					
FMM_IV Comp. 1 Predicted Probability Year 1				0.174*		
				(0.0949)		
FMM IV Comp 1 Predicted Probability Year 2			0.114			
			(0.105)			
FMM-IV Comp 1 Predicted Indicator Year 1						0.140
						(0.0853)
FMM-IV Comp 1 Predicted Indicator Year 2					0.0314	
					(0.0975)	
Sample Size	24500	25000	24500	25000	24500	25000

*** p<0.01, ** p<0.05, * p<0.1, all models include experience, education, race, census region, city size and year.

Table 13: Measurement Error Models using Predicted DER Earnings from FMM

	Men		Women	
	Log CPS Earnings Year 1	Log CPS Earnings Year 2	Log CPS Earnings Year 1	Log CPS Earnings Year 2
FMM Predicted Log DER Earnings Year 1	0.921***		0.956***	
	(0.00842)		(0.00943)	
FMM Predicted Log DER Earnings Year 2		0.925***		0.964***
		(0.00805)		(0.00908)
Mean Squared Error	0.370	0.320	0.278	0.249
Sample Size Rounded	34000	33500	25000	24500

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 14: Measurement Error Models Using FMM Predicted DER Earnings

	Men				Women			
	Respond Both Years		Switchers		Respond Both Years		Switchers	
	Log CPS Earnings Year 1	Log CPS Earnings Year2	Log CPS Earnings Year 1 Only	Log CPS Earnings Year 2 Only	Log CPS Earnings Year 1	Log CPS Earnings Year2	Log CPS Earnings Year 1 Only	Log CPS Earnings Year 2 Only
FMM Predicted Log DER Earnings Year 1	0.914***		0.967***		0.951***		0.988***	
	(0.00874)		(0.0281)		(0.00962)		(0.0348)	
FMM Predicted Log DER Earnings Year 2		0.917***		0.982***		0.966***		0.943***
		(0.00829)		(0.0292)		(0.00949)		(0.0304)
Mean Squared Error	0.346	0.303	0.537	0.472	0.255	0.243	0.444	0.299
Sample Size (Rounded)	30000	30000	4000	3500	22000	22000	3000	2500

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 15: FMM Measurement Error Models Using FMM Predicted DER Earnigns

	Men				Women			
	Log CPS Earnings Year 1		Log CPS Earnings Year 2		Log CPS Earnings Year 1		Log CPS Earnings Year 2	
	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
FMM Predicted Log DER Earnings Year 1	0.912***	1.240***			0.948***	1.394***		
	(0.00694)	(0.169)			(0.00804)	(0.218)		
FMM Predicted Log DER Earnings Year 2			0.918***	1.078***			0.951***	1.534***
			(0.00712)	(0.128)			(0.00806)	(0.203)
Respondent in year 1								
Respondent in year 2	0.767***		0.672***		0.558***		0.549***	
	(0.109)		(0.117)		(0.153)		(0.168)	
Mean Squared Error	0.464	2.205	0.454	1.748	0.423	1.933	0.415	1.691
Predicted Proportion	0.969	0.031	0.96	0.04	0.974	0.026	0.973	0.027
Sample Size Rounded	34000		33500		25000		24500	

*** p<0.01, ** p<0.05, * p<0.1, all models include indicator for year

Table 16a: Probit Models of Non-Response on Measurement Error - Men

	Men					
	Response Year 1	Response Year 2	Response Year 1	Response Year 2	Response Year 1	Response Year 2
Log DER Earnings Year 1	0.800***		0.781***		0.823***	
	(0.110)		(0.109)		(0.108)	
Log DER EARnings Year 1 squared	-0.0340***		-0.0338***		-0.0358***	
	(0.00546)		(0.00522)		(0.00517)	
Log DER Earnings Year 2		0.713***		0.661***		0.693***
		(0.113)		(0.111)		(0.110)
Log DER EARnings Year 2 squared		-0.0295***		-0.0278***		-0.0293***
		(0.00552)		(0.00528)		(0.00524)
Measurment Error Residuals Year 1		-0.0348				
		(0.0213)				
OLS Residuals Year 2 Squared		-0.0117***				
		(0.00334)				
Measurement Error Residuals Year 2	-0.0338					
	(0.0234)					
OLS Residuals Year 2 Squared	-0.0168***					
	(0.00441)			0.320***		
FMM Comp. 1 Predicted Probability Year 1				(0.0686)		
FMM Comp 1 Predicted Probability Year 2			0.229***			
			(0.0722)			
FMM Comp 1 Predicted Indicator Year 1						0.249***
						(0.0621)
FMM Comp 1 Predicted Indicator Year 2					0.135**	
					(0.0651)	
Sample Size	33500	34000	33500	34000	33500	34000

*** p<0.01, ** p<0.05, * p<0.1, all models include experience, education, race, census region, city size and year.

Table 16b: Probit Models of Non-Response on Measurement Error - Women

	Women					
	Response Year 1	Response Year 2	Response Year 1	Response Year 2	Response Year 1	Response Year 2
Log DER Earnings Year 1	1.074***		0.964***		0.985***	
	(0.198)		(0.198)		(0.194)	
Log DER EARnings Year 1 squared	-0.0493***		-0.0430***		-0.0440***	
	(0.00996)		(0.00966)		(0.00951)	
Log DER Earnings Year 2		0.881***		0.815***		0.826***
		(0.179)		(0.176)		(0.174)
Log DER EARnings Year 2 squared		-0.0404***		-0.0367***		-0.0373***
		(0.00902)		(0.00861)		(0.00851)
OLS Measurment Error Residuals Year 1		0.0143				
		(0.0295)				
OLS Residuals Year 2 Squared		-0.0111**				
		(0.00503)				
OLS Measurement Error Residuals Year 2	0.0357					
	(0.0332)					
OLS Residuals Year 2 Squared	-0.000504					
	(0.00695)					
FMM Comp. 1 Predicted Probability Year 1				0.161*		
				(0.0951)		
FMM Comp 1 Predicted Probability Year 2			0.115			
			(0.104)			
FMM Comp 1 Predicted Indicator Year 1						0.156*
						(0.0850)
FMM Comp 1 Predicted Indicator Year 2					0.0900	
					(0.0951)	
Sample Size	24500	25000	24500	25000	24500	25000

*** p<0.01, ** p<0.05, * p<0.1, all models include experience, education, race, census region, city size and year.

Appendix Table 1: DER Earnings standard wage regressions

	Men		Women	
	Year 1	Year 2	Year 1	Year 2
exp	0.0529***	0.0460***	0.0365***	0.0324***
	(0.00135)	(0.00141)	(0.00127)	(0.00130)
exp2	-0.000937***	-0.000806***	-0.000590***	-0.000523***
	(2.86e-05)	(2.88e-05)	(2.71e-05)	(2.68e-05)
black	-0.257***	-0.251***	-0.127***	-0.126***
	(0.0140)	(0.0141)	(0.0116)	(0.0115)
hispanic	-0.210***	-0.210***	-0.146***	-0.137***
	(0.0130)	(0.0131)	(0.0129)	(0.0129)
asian	-0.258***	-0.246***	-0.106***	-0.108***
	(0.0172)	(0.0174)	(0.0169)	(0.0167)
other	-0.139***	-0.147***	-0.00647	-0.000983
	(0.0233)	(0.0234)	(0.0220)	(0.0218)
edELEM	-0.224***	-0.193***	-0.390***	-0.435***
	(0.0279)	(0.0284)	(0.0324)	(0.0325)
ed9th	-0.212***	-0.208***	-0.349***	-0.315***
	(0.0372)	(0.0386)	(0.0475)	(0.0482)
ed10th	-0.178***	-0.144***	-0.279***	-0.316***
	(0.0325)	(0.0331)	(0.0395)	(0.0409)
ed11th	-0.218***	-0.187***	-0.218***	-0.190***
	(0.0287)	(0.0293)	(0.0339)	(0.0339)
ed12nodip	-0.0718**	-0.0714*	-0.129***	-0.112***
	(0.0356)	(0.0377)	(0.0398)	(0.0433)
edGED	-0.103***	-0.0791***	-0.0724***	-0.0764***
	(0.0238)	(0.0276)	(0.0267)	(0.0296)
edsomecoll	0.123***	0.135***	0.124***	0.130***
	(0.0105)	(0.0105)	(0.0104)	(0.0104)
edassoc	0.184***	0.223***	0.236***	0.240***
	(0.0123)	(0.0123)	(0.0117)	(0.0116)
edBA	0.478***	0.479***	0.500***	0.502***
	(0.00969)	(0.00982)	(0.00993)	(0.00992)
edMA	0.700***	0.720***	0.710***	0.716***
	(0.0131)	(0.0131)	(0.0128)	(0.0126)
edPro	1.206***	1.226***	1.086***	1.119***
	(0.0216)	(0.0219)	(0.0272)	(0.0272)
edPhd	0.933***	0.932***	1.029***	1.016***
	(0.0236)	(0.0234)	(0.0295)	(0.0279)
msz_lt100	0.101***	0.111***	0.105***	0.116***
	(0.0154)	(0.0155)	(0.0157)	(0.0156)
msz_100_249	0.114***	0.111***	0.150***	0.143***
	(0.0151)	(0.0152)	(0.0153)	(0.0152)

msz_250_499	0.183***	0.186***	0.224***	0.218***
	(0.0158)	(0.0160)	(0.0161)	(0.0159)
msz_500_999	0.232***	0.235***	0.223***	0.216***
	(0.0162)	(0.0163)	(0.0164)	(0.0162)
msz_1m_249m	0.258***	0.263***	0.311***	0.301***
	(0.0149)	(0.0150)	(0.0151)	(0.0150)
msz_250m_499m	0.380***	0.384***	0.411***	0.411***
	(0.0150)	(0.0152)	(0.0154)	(0.0152)
msz_gt500m	0.317***	0.323***	0.380***	0.382***
	(0.0165)	(0.0167)	(0.0168)	(0.0166)
div2_midatl	-0.0183	-0.0132	-0.0234	-0.0122
	(0.0160)	(0.0162)	(0.0163)	(0.0161)
div3_eastncent	-0.0570***	-0.0683***	-0.0609***	-0.0632***
	(0.0139)	(0.0140)	(0.0143)	(0.0142)
div4_westncent	-0.126***	-0.126***	-0.111***	-0.115***
	(0.0137)	(0.0139)	(0.0140)	(0.0139)
div5_southatl	-0.0882***	-0.0874***	-0.0836***	-0.0743***
	(0.0131)	(0.0132)	(0.0133)	(0.0132)
div6_eastscent	-0.117***	-0.121***	-0.134***	-0.134***
	(0.0193)	(0.0195)	(0.0189)	(0.0187)
div7_westscent	-0.0951***	-0.0744***	-0.118***	-0.116***
	(0.0158)	(0.0159)	(0.0159)	(0.0158)
div8_mtn	-0.0347**	-0.0369**	-0.0361**	-0.0375**
	(0.0146)	(0.0147)	(0.0154)	(0.0152)
div9_pacific	0.0120	0.0145	0.0319**	0.0412***
	(0.0141)	(0.0143)	(0.0147)	(0.0145)
yr2006	-0.0164	0.00331	-0.00380	-0.00706
	(0.0103)	(0.0109)	(0.0105)	(0.0108)
yr2007	0.0245**	-0.0275**	0.0127	-0.0143
	(0.0104)	(0.0108)	(0.0106)	(0.0107)
yr2008	0.0276***	-0.00578	0.00811	-0.0181*
	(0.0106)	(0.0109)	(0.0106)	(0.0107)
yr2009	0.0158	-0.00142	0.0285***	-0.00126
	(0.0108)	(0.0111)	(0.0109)	(0.0107)
Constant	9.876***	9.957***	9.658***	9.738***
	(0.0225)	(0.0238)	(0.0228)	(0.0234)
Sample Size	41000	41000	30000	30000

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 2: DER Earnings regression component one results

	Men		Women	
	Year 1	Year 2	Year 1	Year 2
Experience	0.111***	0.0842***	0.0916***	0.0640***
	(0.00624)	(0.00706)	(0.00584)	(0.00647)
Experience squared	-0.00455***	-0.00266***	-0.00451***	-0.00249***
	(0.000500)	(0.000534)	(0.000478)	(0.000502)
black	-0.260***	-0.259***	-0.128***	-0.126***
	(0.0113)	(0.0113)	(0.00945)	(0.00935)
hispanic	-0.203***	-0.198***	-0.142***	-0.137***
	(0.0108)	(0.0108)	(0.0108)	(0.0106)
asian	-0.199***	-0.197***	-0.0596***	-0.0620***
	(0.0153)	(0.0150)	(0.0146)	(0.0144)
other	-0.142***	-0.151***	-0.0236	-0.0223
	(0.0187)	(0.0188)	(0.0176)	(0.0174)
edELEM	-0.259***	-0.260***	-0.395***	-0.377***
	(0.0236)	(0.0234)	(0.0264)	(0.0265)
ed9th	-0.234***	-0.220***	-0.363***	-0.350***
	(0.0308)	(0.0317)	(0.0383)	(0.0385)
ed10th	-0.180***	-0.189***	-0.278***	-0.292***
	(0.0263)	(0.0263)	(0.0313)	(0.0331)
ed11th	-0.211***	-0.192***	-0.200***	-0.209***
	(0.0234)	(0.0237)	(0.0278)	(0.0273)
ed12nodip	-0.0867***	-0.0881***	-0.145***	-0.110***
	(0.0289)	(0.0305)	(0.0331)	(0.0357)
edGED	-0.0908***	-0.0923***	-0.0663***	-0.0835***
	(0.0191)	(0.0218)	(0.0218)	(0.0244)
edsomecoll	0.116***	0.124***	0.129***	0.131***
	(0.00853)	(0.00847)	(0.00855)	(0.00850)
edassoc	0.189***	0.215***	0.250***	0.255***
	(0.00992)	(0.00982)	(0.00960)	(0.00948)
edBA	0.475***	0.480***	0.510***	0.511***
	(0.00809)	(0.00811)	(0.00821)	(0.00819)
edMA	0.664***	0.684***	0.696***	0.708***
	(0.0109)	(0.0108)	(0.0103)	(0.0102)
edPro	1.219***	1.239***	1.105***	1.118***
	(0.0205)	(0.0205)	(0.0243)	(0.0244)
edPhd	0.915***	0.923***	1.021***	1.022***
	(0.0204)	(0.0199)	(0.0244)	(0.0234)
msz_lt100	0.0611***	0.0670***	0.0912***	0.0897***
	(0.0125)	(0.0125)	(0.0127)	(0.0126)
msz_100_249	0.0747***	0.0826***	0.127***	0.122***
	(0.0123)	(0.0123)	(0.0124)	(0.0124)
msz_250_499	0.137***	0.138***	0.202***	0.190***
	(0.0130)	(0.0130)	(0.0131)	(0.0130)
msz_500_999	0.185***	0.192***	0.207***	0.204***
	(0.0133)	(0.0133)	(0.0134)	(0.0133)
msz_1m_249m	0.220***	0.217***	0.281***	0.271***

	(0.0122)	(0.0121)	(0.0123)	(0.0122)
msz_250m_499m	0.339***	0.349***	0.405***	0.396***
	(0.0124)	(0.0124)	(0.0126)	(0.0125)
msz_gt500m	0.311***	0.309***	0.390***	0.374***
	(0.0138)	(0.0138)	(0.0138)	(0.0137)
div2_midatl	-0.00556	0.00865	-0.0330**	-0.0319**
	(0.0133)	(0.0133)	(0.0134)	(0.0133)
div3_eastncent	-0.0444***	-0.0509***	-0.0676***	-0.0773***
	(0.0114)	(0.0113)	(0.0116)	(0.0115)
div4_westncent	-0.118***	-0.112***	-0.111***	-0.115***
	(0.0112)	(0.0112)	(0.0114)	(0.0113)
div5_southatl	-0.0495***	-0.0498***	-0.0757***	-0.0742***
	(0.0108)	(0.0108)	(0.0109)	(0.0109)
div6_eastscent	-0.112***	-0.110***	-0.152***	-0.157***
	(0.0158)	(0.0158)	(0.0153)	(0.0151)
div7_westscent	-0.0663***	-0.0577***	-0.137***	-0.135***
	(0.0132)	(0.0131)	(0.0130)	(0.0129)
div8_mtn	-0.0142	-0.00172	-0.0375***	-0.0319**
	(0.0121)	(0.0121)	(0.0126)	(0.0126)
div9_pacific	0.0382***	0.0462***	0.0364***	0.0397***
	(0.0117)	(0.0117)	(0.0121)	(0.0120)
yr2006	-0.00693	0.00435	-0.00733	-0.0114
	(0.00866)	(0.00908)	(0.00880)	(0.00870)
yr2007	0.0115	-0.00381	7.81e-05	-0.0184**
	(0.00871)	(0.00905)	(0.00881)	(0.00872)
yr2008	0.0150*	0.00160	-0.00181	-0.00324
	(0.00891)	(0.00905)	(0.00887)	(0.00878)
yr2009	0.00319	0.0103	0.0233***	-0.00290
	(0.00908)	(0.00922)	(0.00896)	(0.00892)
Constant	9.686***	9.795***	9.514***	9.653***
	(0.0287)	(0.0336)	(0.0271)	(0.0306)
Probability of Component 1	0.902	0.902	0.924	0.934
Sample Size	41000	41000	30000	30000

*** p<0.01, ** p<0.05, * p<0.1