Consider the Source: Using Administrative Records to Estimate Returns to Earnings

Michael Gideon, Amazon (formerly U.S. Census Bureau)

Misty L. Heggeness, U.S. Census Bureau

Marta Murray-Close, U.S. Census Bureau

Samuel L. Myers, Jr., University of Minnesota

September 18, 2017

This paper was prepared for presentation at the Association for Public Policy Analysis and Management (APPAM) 2017 Fall Research Conference. It was developed to promote research and advancements in our understanding of poverty measurement. In that spirit and to encourage discussion and thoughtful feedback at early stages of our work, this paper has undergone a more limited review than official Census Bureau reports. All views and any errors are solely those of the authors and do not necessarily reflect any official position of the Bureau. Do not cite or distribute without author permission.

ABSTRACT

In this paper, we explore how estimated returns to human capital characteristics in predicting earnings are affected by measurement error. We find that estimated returns to age and education may depend on the source data for earnings and that the impact of changing the source data may differ by race. Using standard analytical tools for measuring wage discrimination, we find that less of the racial wage gap is explained by worker characteristics when using a Current Population Survey (CPS) self-reported earnings measure rather than a measure of earnings from the Social Security Administration's Detailed Earnings Record (DER). While other studies have examined issues of misreporting, specifically at the tail ends of the income distribution, our analysis extends beyond distributional changes to examine the impact of the source of earnings data on inequality measurement. Overall, our results provide informative and relevant information for understanding the extent to which self-reported earnings data and reported earnings from administrative records influences the estimation of inequality and our interpretation of factors that drive earnings inequality by race, specifically the black-white earnings gap.

INTRODUCTION

Economic research on earnings and income relies primarily on self-reported earnings from household surveys like the Current Population Survey (CPS).[1] One area, which has relied primarily, if not solely, on self-reported earnings, is in the study of wage and earnings inequality. Federal agencies calculate Gini coefficients and other inequality measures using the CPS and other household surveys (Semega, Fontenot, and Kollar 2017; Posey 2016). Other investigative studies as early as the 1970s have focused on analyzing black-white earnings gaps with the CPS and related household surveys (Masters 1974; Smith and Welch 1977; Darity et al. 1998; O'Neill 1990; Arias et al. 2004; Gabriel 2004; Cunningham and Jacobsen 2008).

In this paper, we examine the effect, if any, of relying on self-reported household survey earnings data on inequality measurement. We also study the intersection of demographics and earnings measurement error, specifically the black-white earnings gap. Our study is twofold. First, we explore differences in earnings along age and educational attainment thresholds by two types of source data: the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the Social Security Administration's Detailed Earnings Record (DER). To the extent that individuals self-report earnings inaccurately, with less precision, or to the extent that top coding influences inequality measures calculated using the CPS ASEC, we capture the overall magnitude of these effects for blacks, whites, and the black-white differential. Second, we examine the extent to which differences in earnings between CPS ASEC self-reported earnings and administrative records of earnings from the Social Security Administration influence commonly used analytical tools for wage discrimination and earnings inequality. Overall, this paper contributes to the literature by expanding our understanding of potential issues to be aware of when using household survey earnings data to study and measure inequality,

[1] For more information on the Current Population Survey, see: https://www.census.gov/programs-surveys/cps.html or https://cps.ipums.org/cps/.

particularly differences in earnings across demographic groups.

BACKGROUND

For years, researchers have studied black-white earnings inequality within the United States. There are many ways to measure intergroup earnings inequality, and different measures may give very different pictures of relative well-being across groups. Historically, the most common has been to focus on differences in mean or median earnings between groups (Masters 1974; Smith and Welch 1977; Darity et al. 1998; O'Neill 1990; Gabriel 2004). More recent studies expand the focus to distributional parameters beyond the mean (Bayer and Charles 2016).

While there has been some discussion of the impact of the quality of earnings data on poverty measurement (Hokayem et al. 2015), there is very little discussion in the literature of the quality of earnings data and the potential for erroneous estimation of inequality measures due to confounding issues in data collection and reporting. This is surprising in light of stylized facts about how data quality varies across the distribution of earnings. First, self-reported earnings are "mean-reverting"—respondents at the low end of the earnings distribution over report their earnings and those at the top of the distribution underreport their earnings (e.g. Bollinger 1998; Bound et al 2001). Second, non-response exhibits "trouble in the tails" (Hokayem et al. 2015, 2016; Bollinger et al. 2015)—respondents with the lowest earnings in tax records are least likely to respond to the survey and answer questions about earnings. As highlighted by Meyer et al. (2015), survey non-response rates have increased substantially over the past couple decades. Hokayem et al. (2015) show that non-response affects measures of poverty, suggesting that imputations could also bias estimates of inequality. Finally, non-response rates differ by race (Gideon et al. 2017).

In light of these stylized facts, how does source data affect national statistics on racial earnings inequality and returns to investments in human capital? In prior work (Gideon et al. 2017), we examined the role of source data and measurement error on the black-white earnings gap. In this paper, we extend that analysis by studying the way in which non-response and differences across the earnings distribution interact with age and educational attainment to influence the black-white earnings gap. We highlight the importance of accounting for differences in earnings by comparing the CPS ASEC, a monthly CPS supplement capturing detailed information on employment, earnings, and income, and Social Security Administration's Detailed Earnings Record (DER) measures across the earnings distribution, as well as differences in the distributions by race.

DATA

Our data stems from two sources: the CPS ASEC and the DER file.[2] Each year, between February and April, the CPS surveys approximately 99,000 households about their incomes and characteristics of their employment in the previous calendar year. We use data from CPS for 2005-2013, corresponding to tax years and DER earnings data for 2004-2012, and link to administrative earnings records for the same years. Years prior to 2004 are limited due to geography and concerns about the quality of linkages to the administrative records, and years after 2012 are limited by the availability of tax earnings data. We include demographic and socioeconomic variables from the CPS, as well as a measure of earnings from the CPS and a measure of earnings from the DER. We use information from the CPS on wage and salary earnings and the characteristics of the longest-held job, particularly information used to determine full-time full-year work status and the class of employer (private, public, or self-employed).

[2] Differences are often attributed to measurement error in self-reported CPS earnings, although more recent work emphasized other explanations for differences between self-reported and administrative earnings (Abowd and Stinson 2013; Kapteyn and Ypma 2007). We remain agnostic about whether data from either source are "true," focusing instead on how differences between the two measures might affect earnings gaps estimated using these measures.

We construct two individual-level measures of total yearly earnings, one based on survey responses in the CPS and one based on administrative records in the DER. Our CPS measure includes all wage and salary earnings because the CPS does not collect information about every job separately. Administrative earnings data come from the DER.[3] We define annual earnings for each job in the DER as the sum of Box 1 (wages, salaries, bonuses, etc.) and Box 12 amounts (tax-deferred contributions to employer-sponsored retirement plans, such as 401(k)s) from W-2 Forms. Annual earnings for each job are summed together to get total yearly earnings. Both earnings variables are adjusted to 2015 dollars using the Consumer Price Index (CPI).

The baseline sample consists of men who were between the ages of 25 and 64 at the time of their CPS interview, who identified their race and ethnicity as non-Hispanic white or non-Hispanic black, and whose longest-held job during the survey reference year was in a non-agricultural, non-military occupation.[4] To maximize the comparability of our CPS earnings measure with our DER earnings measure, which does not include earnings from self-employment, we restrict the sample to men whose longest-held job was with a public or private-sector employer. Finally, we restrict the sample to individuals with strictly positive earnings in both the CPS and the DER in a given year. This leaves us with a sample of 340,004 person-year observations.

All observations have strictly positive earnings in the CPS, and we do not have cases in which people report zero earnings in CPS but have earnings in the DER. However, there are two reasons a person in the CPS might not match to the DER. Most importantly, he or she might not have been assigned a unique personal identification key (PIK). The PIK is a unique identifier developed by the

[3] The DER contains information on all jobs with W-2 forms, as well as self-employment information from tax returns. We focus on private- and public-sector workers, and therefore focus on earnings in the DER that come from W-2 records.
[4] Less than 1.5 percent of the baseline sample reported multiple racial identifications in the CPS. These people were assigned to a single-race category using a standard recoded race and ethnicity variable. When revising this work, we plan to restrict our estimation samples to people who reported only one racial identification.

Census Bureau that facilitates linking individuals to other individual-level datasets.[5] Respondents who do not have validated PIKs are removed and observations are re-weighted to account for the probability of being assigned a PIK. Second, people can have a validated PIK, report working in the CPS, but not have earnings reports in the DER. This problem arises when there are informal jobs that do not generate W-2 forms. For our analysis on the black-white earnings gap, we need positive earnings data from both data sources. Since we take only individuals with positive, non-imputed earnings in both sources of data, we exclude individuals with reported informal earnings that are not recorded in the DER.[6]

Respondents may refuse to answer specific questions in the CPS or refuse to answer the entire Annual Social and Economic Supplement survey. We remove observations that are imputed due to supplement nonresponse and observations that have imputed values for earnings from their longest held job.[7] Because blacks and whites have different earnings distributions and imputation rates, nonresponse (and errors due to imputation) could explain differences in the black-white earnings gap when using earnings measures from CPS versus the DER. There is, however, a cost to excluding the imputations. Unless the underlying data are missing at random, observations with full-record or earnings imputations are likely to differ from those without imputations. To mitigate such concerns, we adjust weights to generate population-level estimates.

To obtain nationally representative results, we compute point estimates using CPS person weights. Standard errors and corresponding confidence intervals are computed using CPS replicate weights. Both the person weights and replicate weights are adjusted for the probability that a CPS respondent is successfully matched with an administrative earnings record.

[5] See Wagner and Layne (2014) for more information on the PIK assignment process.

[6] If informal arrangements vary across time differently by race, then changes in the difference in the black-white earnings gaps using CPS and DER over time could be impacted by the changes in the samples.

[7] It's possible that people report their earnings from their main job but have imputed values for their other earnings. These cases are treated as not being imputed.

METHODS AND RESULTS

In this section we describe the methods used to estimate earnings inequality using two different sources of data for the measurement of earnings. We also examine the relevance of differences in inequality and returns to human capital investments by source data using standard analytical tools and compare our results.

The difference between CPS and DER earnings varies with worker characteristics

Table 1 shows the mean, among black and white men in each of four ten-year age groups (25-34, 35-44, 45-54, and 55-64), of the difference between the log of the earnings reported in the CPS survey and the log of the earnings recorded in the DER in the same reference year (we refer to this difference as the "reporting gap"). Columns 1 and 2 of Table 1 show the mean reporting gaps for white and black men, respectively. Column 3 shows the difference between the mean reporting gaps for black and white men.

Among white men, the mean reporting gap is positive and statistically significant in each age group. The point estimates decline monotonically across age groups, and the means for the 35-44, 45-54, and 55-64 age groups all differ significantly from the mean for the 25-34 age group. This pattern of over reporting compared to the DER declines with age suggesting that the age-earnings profiles of white men are flatter (earnings appear to grow more slowly with age) when earnings are measured with survey data rather than administrative records.

Among black men, the mean reporting gap is positive and statistically significant for the 25-34 and 35-44 age groups, statistically indistinguishable from zero for the 45-54 age group, and negative and statistically significant for the 55-64 age group. Again, the point estimates decline monotonically across age groups, and the means for the 35-44, 45-54, and 55-64 age groups all differ significantly from the mean for the 25-34 age group. As with white men, this pattern of reporting gaps declining with age suggests that age-earnings profiles of black men are flatter when earnings are measured with survey data from the CPS.

When we compare the mean reporting gaps of black and white men in the same age groups, we find that the flattening of the age-earnings profile in the CPS data compared with the DER data is more pronounced for black men than white men. Among the youngest men (the 25-34 age group), the mean reporting gap for black men is statistically larger than the mean reporting gap for white men; among the oldest men (the 45-54 and 55-64 age groups), it is statistically smaller. These results suggest that the earnings returns to age may appear lower for black men relative to white men when earnings are measured with the CPS data rather than the DER data.

Table 2 shows the mean reporting gaps for black and white men and the difference between these gaps, at four levels of education. Among white men, the mean reporting gap does not vary much with education. The gaps for white men with a high school degree, some college, and a college degree or more are not statistically different from the gap for men with less than a high school degree. Among black men, in contrast, the mean reporting gap increases with education: the gaps for men with some college and a college degree or more are statistically larger than the gap for men with less than a high school degree.

Comparing the mean reporting gaps for black and white men at each level of education, we find that the gap for black men is smaller than the gap for white men at the lowest level of education but that the reverse is true at the highest level of education. These results suggest that the earnings returns to education may appear higher for black men relative to white men when earnings are measured with the CPS data rather than the DER data.

The estimated returns to worker characteristics depend on the measure of earnings

Taken together, the results in Tables 1 and 2 show that the choice of source data for earnings may influence the measured returns to worker characteristics like age and education differentially by race. In particular, the results suggest that using survey data from the CPS rather than administrative records from the DER may decrease the measured return to age and increase the measured return to education for black men relative to white men. Table 3 shows that the same patterns hold in the context of the following ordinary least squares (OLS) earnings regression:

$$(1)\ \log(earnings) = \alpha + \beta_1(Age_{35-44}) + \beta_2(Age_{45-54}) + \beta_3(Age_{55-64}) + \gamma_1(High\ School) + \gamma_2(Some\ College) + \gamma_3(College\ Degree\ or\ Higher)$$

Column 1 of Table 3 shows the difference in the coefficients from estimating equation (1) on the sample of white men using the CPS versus the DER earnings measures. Column 2 shows the corresponding results for black men, and Column 3 compares the gaps between the CPS and DER coefficients for black men with the gaps for white men. The results in Table 3 show that the gains from being in one of the oldest age groups (45-54 or 55-64) rather than the youngest (25-34) are statistically lower for black men relative to white men when earnings are measured using the CPS rather than the DER. The gains from having one of the highest levels of education (some college or a college degree or more) rather than the lowest (less than a high school degree) are statistically higher for black men relative to white men when using the CPS. Thus, with regression coefficients as with means, the earnings returns to age appear lower and the earnings returns to education higher, for black men relative to white men, when earnings are measured with the CPS data rather than the DER data.

Extension: Oaxaca-Blinder decompositions with CPS and DER earnings measures

We have seen that the estimated returns to age and education may depend on the source data for earnings and that the impact of changing the source data may differ by race. This finding has implications for research focused on the returns to worker characteristics. More broadly, the finding suggests that the choice of source data may matter for research that uses estimates of returns to worker characteristics to answer other questions.

One example of a research method that uses estimated returns to worker characteristics is the Oaxaca-Blinder decomposition, which decomposes the difference in earnings between two groups (for example, black and white men) into a component that is accounted for by differences in earnings-relevant characteristics (for example, age and education) between the groups, a component that is accounted for by differences in returns to those characteristics, and (in some versions of the decomposition) a component that is accounted for by the interaction between differences in characteristics and differences in returns (Jann 2008). To illustrate the impact the choice of source data may have on estimates of these components, we decompose the log-earnings gap between black and white men using the following Oaxaca-Blinder equation, first with the CPS earnings measure and then with the DER earnings measure.[8]

$$\text{(2)} \qquad \overline{Y}_W - \overline{Y}_B = \left(\overline{X}_W - \overline{X}_B\right)\beta_B + \overline{X}_B(\beta_W - \beta_B) + \left(\overline{X}_W - \overline{X}_B\right)(\beta_W - \beta_B)$$

$\overline{Y}_W$ and $\overline{Y}_B$ are the average values of log-earnings for white and black men. $\overline{X}_W$ and $\overline{X}_B$ are vectors containing the mean values of the regressors in equation (1), and β_W and β_B are vectors of

[8] In the discussion that follows, we do not test differences between the CPS and DER decompositions for statistical significance. Our aim is not to make statistically valid claims about the real-world causes of racial earnings disparities but simply to illustrate the potential implications of measurement choices for a widely-used econometric method.

coefficients obtained from estimating equation (1) on the samples of white and black men.

The first term on the right-hand side of equation (2) is the explained component of the racial earnings gap. Intuitively, it answers this question: Given their actual returns to characteristics, how much more would black men earn if they had the characteristics of white men? The second term on the right-hand side of the equation is the unexplained component. It answers this question: Given their actual characteristics, how much more would black men earn if they had the same returns to characteristics as white men (that is, if they were "treated as" white men)? The final term on the right-hand size of equation (2) is the interaction. It answers this question: Beyond the amounts captured by the first two components, how much more would black men earn if they had the characteristics of white men *and* were treated as white men?

Table 4 compares the results from estimating equation (2) with the CPS and DER earnings measures. Because the explained component of the racial earnings gap in equation (1) weights racial differences in characteristics by the actual black returns to characteristics, we would expect the CPS-DER gaps in black returns (Column 2 of Table 3) to translate to CPS-DER gaps in the explained component of the decomposition. For example, given that the black men in our estimation sample are younger than the white men (with some abuse of notation, $\overline{X}_W - \overline{X}_B > 0$), we would expect the negative CPS-DER gap in black returns to age ($\beta_{B,CPS} < \beta_{B,DER}$) to reduce the portion of the racial earnings gap accounted for by racial differences in age. This is indeed what we observe – the "endowments" effect of age in Table 4 is 0.023 when using the DER earnings measure but just 0.015 when using the CPS earnings measure.

Because the unexplained component of the racial earnings gap in equation (1) depends on racial differences in returns to characteristics, we would expect the racial differences in the CPS-DER gaps in returns (Column 3 of Table 3) to translate to CPS-DER gaps in the unexplained component of the decomposition. For example, because black men have lower returns to age relative to white men when earnings are measured with the CPS rather than the DER (again with some abuse of notation, $\beta_{W,CPS}$ - $\beta_{B,CPS}$ > $\beta_{W,DER}$ - $\beta_{B,DER}$), we would expect the change in black earnings from treating black men's age like white men's to be more positive (or less negative) in the CPS decomposition than the DER decomposition. Again, this is what we observe – the "coefficients" effect of age in Table 4 is -0.056 when using the DER earnings measure and -0.028 when using the CPS.

The extent to which the racial wage gap is explained by worker characteristics when using the CPS earnings measure rather than the DER earnings measure depends on the impact of the source data on the portion explained by each of the characteristics. In our illustration, the source of the data impacts the portion explained by both age and education. We have seen that the portion of the racial wage gap explained by age is smaller when using the CPS earnings measure. On the other hand, Table 4 shows that the portion explained by education is slightly larger. As it turns out, the impact of the source data on age dominates, so less of the racial wage gap is explained by worker characteristics when using the CPS earnings measure rather than the DER earnings measure.

CONCLUSION

We find that not only are inequality measures sensitive to source data and non-response, but so are other analytical tools used to study inequality. In particular, our results show that returns to earnings by age and education vary by source data for both blacks and whites. In particular, our results provide evidence that when using CPS self-reported earnings rather than the DER administrative earnings records:

- age-earnings profiles of both white and black men are flatter,
- the earnings returns to age may appear lower for black men relative to white men,

- the earnings returns to education may appear higher for black men relative to white men, and
- differences in the factors driving earnings inequality and the magnitude of those factors exist, resulting in a cautionary note to those using CPS self-reported earnings to measure earnings returns in investments in human capital by race.

These finding are, to our knowledge, the first exposition providing evidence that not only does nonresponse and measurement error directly influence earnings estimations across the distribution by race, but it also affects earnings returns to human capital and predicting those returns.

Our results illustrate how source data can impact estimates of national statistics of inequality and measures of returns to human capital investments. In particular, we show that differences between CPS and DER earnings data influences not only the direct estimation of black-white earnings gaps, but also the estimations of earnings returns to age and education by race. Future studies examining racial earnings inequality, non-response, and source data should consider the broader implications of data quality on demographic characteristics and their interpretations in general.

Table 1. Mean difference between earnings as reported in survey data (CPS) and earnings as recorded in administrative data (DER), by race and age group

Age group		White CPS - DER	Black CPS – DER	(Black CPS – DER) – (White CPS – DER)
Age 25-34	Mean	0.0910*	0.1168*	0.0258*
	Standard error	-0.0025	-0.0079	-0.0083
	Observations	10,143	3,042	13,185
	p (= Age 25-34)	--	--	--
Age 35-44	Mean	0.0577*	0.0666*	0.0089
	Standard error	-0.0024	-0.0073	-0.0076
	Observations	78,529	12,387	90,916
	p (= Age 25-34)	[0.0000]	[0.0000]	[0.1356]
Age 45-54	Mean	0.0259*	0.0074	-0.0185*
	Standard error	-0.0022	-0.0063	-0.0067
	Observations	90,779	13,503	104,282
	p (= Age 25-34)	[0.0000]	[0.0000]	[0.0000]
Age 55-64	Mean	0.0249*	-0.0299*	-0.0548*
	Standard error	-0.0029	-0.0077	-0.0082
	Observations	120,604	11,017	131,621
	p (= Age 25-34)	[0.0000]	[0.0000]	[0.0000]

Source: Survey data are from the 2005-13 Annual Social and Economic Supplement of the Current Population Survey (CPS). Administrative data are from the 2004-2012 Detailed Earnings Records (DER) from the Social Security Administration.

*Notes: * indicates that the mean or difference in means is statistically different from zero at the 5-percent level. The numbers in square brackets are p-values from tests of the null hypothesis that the mean for the specified age group is equal to the mean for the 25-34 age group. The estimation sample includes men who reported in the CPS that they were single-race black or white non-Hispanic, aged 25 to 64, held a public-or private-sector job, and had positive wage or salary earnings, excluding those who reported that they were self-employed, worked in agriculture, or were members of the armed forces. The estimation sample excludes men who met these criteria if their CPS earnings were imputed or they lacked a linked DER record with positive earnings. Means were estimated using the CPS person weights multiplied by the estimated inverse probability of having reported (non-imputed) CPS earnings and a linked DER record with positive earnings.*

Table 2. Mean difference between log earnings as reported in survey data (CPS) and log earnings as recorded in administrative data (DER), by race and educational attainment

Highest degree		White CPS - DER	Black CPS – DER	(White CPS – DER) – (Black CPS – DER)
Less than high school	Mean	0.0543*	0.014	-0.0403*
	Standard error	-0.0074	-0.0156	-0.0172
	Observations	10,143	3,042	13,185
	p (= Less than high school)	--	--	--
High school	Mean	0.0492*	0.0441*	-0.0051
	Standard error	-0.0023	-0.0076	-0.008
	Observations	78,529	12,387	90,916
	p (= Less than high school)	[0.5084]	[0.0822]	[0.0633]
Some college	Mean	0.0597*	0.0646*	0.0048
	Standard error	-0.0022	-0.0058	-0.0062
	Observations	90,779	13,503	104,282
	p (= Less than high school)	[0.4767]	[0.0024]	[0.0139]
College degree or more	Mean	0.0443*	0.0572*	0.0129*
	Standard error	-0.002	-0.006	-0.0063
	Observations	120,604	11,017	131,621
	p (= Less than high school)	[0.1939]	[0.0097]	[0.0038]

Source: Survey data are from the 2005-13 Annual Social and Economic Supplement of the Current Population Survey (CPS). Administrative data are from the 2004-2012 Detailed Earnings Records (DER) from the Social Security Administration.

*Notes: * indicates that the mean or difference in means is statistically different from zero at the 5-percent level. The numbers in square brackets are p-values from tests of the null hypothesis that the mean for the specified education group is equal to the mean for the group with less than a high school degree. The estimation sample and weights are described in the notes to Table 1.*

Table 3. Difference between coefficients from log-earnings regressions using earnings as reported in survey data (CPS) and earnings as recorded in administrative data (DER)

	White CPS – DER	Black CPS – DER	(Black CPS – DER) – (White CPS – DER)
High school	-0.0035	0.0227	0.0263
	(0.0077)	(0.0171)	(0.0188)
Some college	0.0045	0.0405*	0.0360*
	(0.0077)	(0.0165)	(0.0182)
College or more	-0.0127	0.0327*	0.0453*
	(0.0076)	(0.0166)	(0.0183)
Age 35-44	-0.0335*	-0.0500*	-0.0165
	(0.0034)	(0.0108)	(0.0113)
Age 45-54	-0.0657*	-0.1086*	-0.0428*
	(0.0033)	(0.0102)	(0.0107)
Age 55-64	-0.0665*	-0.1447*	-0.0782*
	(0.0038)	(0.0111)	(0.0118)
Observations	600,110	600,110	680,008

Source: Survey data are from the 2005-13 Annual Social and Economic Supplement of the Current Population Survey (CPS). Administrative data are from the 2004-2012 Detailed Earnings Records (DER) from the Social Security Administration.

*Notes: * indicates that the coefficient or difference in coefficients is statistically different from zero at the 5-percent level. The estimation sample and weights are described in the notes to Table 1.*

Table 4. Oaxaca-Blinder decomposition of black-white log-earnings gap using earnings as reported in survey data (CPS) and earnings as recorded in administrative data (DER)

	DER	CPS
Overall		
White log(Earnings)	10.592*	10.642*
	(0.002)	(0.002)
Black log(Earnings)	10.255*	10.306*
	(0.006)	(0.006)
Difference in log(Earnings)	0.337*	0.336*
	(0.007)	(0.006)
Endowments	0.130*	0.124*
	(0.003)	(0.003)
Coefficients	0.226*	0.231*
	(0.006)	(0.006)
Interaction	-0.019*	-0.018*
	(0.002)	(0.002)
Endowments		
Age	0.023*	0.015*
	(0.002)	(0.001)
Education	0.107*	0.108*
	(0.003)	(0.003)
Coefficients		
Age	-0.056*	-0.028*
	(0.010)	(0.009)
Education	-0.122*	-0.154*
	(0.024)	(0.023)
Intercept	0.404*	0.412*
	(0.029)	(0.027)
Interaction		
Age	-0.008*	-0.004*
	(0.001)	(0.001)
Education	-0.011*	-0.015*
	(0.002)	(0.002)
Observations	340,004	340,004

Source: Survey data are from the 2005-13 Annual Social and Economic Supplement of the Current Population Survey (CPS). Administrative data are from the 2004-2012 Detailed Earnings Records (DER) from the Social Security Administration.

*Notes: * indicates that the coefficient or difference in coefficients is statistically different from zero at the 5-percent level. The estimation sample and weights are described in the notes to Table 1.*

REFERENCES

Abowd, John M., and Martha H. Stinson. 2013. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *The Review of Economics and Statistics* 95:5, 1451-1467.

Arias, Omar, Gustavo Yamada, and Luis Tejerina. 2004. "Education, Family Background, and Racial Earnings Inequality in Brazil." *International Journal of Manpower* 25:3/4, 355.

Bayer, Patrick, and Kerwin Kofi Charles. 2016. "Divergent Paths: Structural Change, Economic Rank, and the Evoluation of Black-White Earnings Differences, 1940-2014." *NBER Working Paper 22797*.

Bollinger, Christopher R. 1998. "Measurement Error in the Current Population Survey: A Nonparametric Look." *Journal of Labor Economics* 16:3, 576-594.

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch." *Unpublished Working Paper*, September 2015.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics* 5, 3705-3843.

Cunningham, Wendy, and Joyce P. Jacobsen. 2008. "Earnings Inequality Within and Across Gender, Racial, and Ethnic Groups in Four Latin American Countries." *World Bank Policy Research Working Paper 4591*.

Darity, William A. Jr., Samuel L. Myers, Jr., and Chanjin Chung. 1998. "Racial Earnings Disparities and Family Structure." *Southern Economic Journal* 65:1, 20-41.

Gabriel, Paul E. 2004. "Differences in Earnings, Skills and Labour Market Experience Among Young Black and White Men." *Applied Economic Letters* 11, 337-341.

Gideon, Michael, Misty L. Heggeness, Marta Murray-Close and Samuel L. Myers, Jr. 2017. "Examining the Black-White Earnings Differential with Administrative Records." SEHSD Working Paper 2017-32.

Hokayem, Charles, Christopher Bollinger, and James P. Ziliak. 2015. "The Role of CPS Nonresponse in the Measurement of Poverty." *Journal of the American Statistical Association* 110:511, 935-945.

Hokayem, Charles, Trivellore Raghunathan, and Jonathan Rothbaum. "Sequential Regression Multivariate Imputation in the Current Population Survey Annual Social and Economic Supplement." *Unpublished Manuscript*, May 2, 2016.

Jann, Ben (2008). The Blinder-Oaxaca Decomposition for Linear Regression Models. *Stata Journal* 8(4): 453-479.

Kapteyn, A., and Ypma, J.Y., "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," Journal of Labor Economics, Vol. 25(3), 2007, 513-551.

Masters, Stanley H. 1974. "The Effect of Educational Differences and Labor-Market Discrimination on the Relative Earnings of Black Males." *The Journal of Human Resources* 9:3, 342-360.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan, 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29:4, 199-226.

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences Between Black and White Men." *Journal of Economic Perspectives* 4:4, 25-45.

Posey, Kirby G. "Household Income: 2015." *U.S. Census Bureau Technical Report ACSBR/15-02*, September 2016.

Semega, Jessica L., Kayla R. Fontenot, and Melissa A. Kollar. "Income and Poverty in the United States: 2016." *U.S. Census Bureau Technical Report P60-259*, September 2017.

Smith, James P. and Finish R. Welch. 1977. "Black-White Male Wage Ratios: 1960-70." *The American Economic Review* 67:3, 323-338.