

Measuring Systematic Wage Misreporting by Socio-demographic Groups

Christian Imboden, University of Oregon
John Voorheis, US Census Bureau
Caroline Weber, University of Oregon

March 7, 2018

This presentation is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

Overview

- ▶ **Measurement error** is an important issue for data providers and data users
 - ▶ Data users: can reduce model efficiency & bias estimates
 - ▶ Data providers: reduces quality of data released
- ▶ Understanding the sources and socio-demographic correlates of measurement can help:
 - ▶ Data users account for measurement error to avoid making misleading inferences and estimating parameters inaccurately
 - ▶ Data providers improve reliability (e.g. improving editing and imputation)

Why Wages?

- ▶ Data use: impacts of **socio-demographic characteristics** on **income** or **wages**
 - ▶ Black-White wage gap
 - ▶ Returns to schooling
- ▶ If measurement error for income is correlated with socio-demographic variables, estimates will be inaccurate
- ▶ We examine sources of **non-zero mean, systematic measurement error** by **demographics** in **survey wage data**

What We Do

- ▶ Validate responses to the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) against administrative wage records
- ▶ We link the CPS ASEC between 2001-2016 with:
 - ▶ Internal Revenue Service (IRS) 1040 tax returns, 2000-2015
 - ▶ Social Security Administration (SSA) Detailed Earnings Record (DER), 2000-2012
 - ▶ IRS W-2s, 2005-2015
- ▶ We rely on the accuracy of SSA/IRS wage data as a benchmark
 - ▶ Interpret differences as misreporting on the CPS

Data

- ▶ Sample restrictions:
 - ▶ Individuals 25-55 with non-zero survey and administrative records wage amounts
 - ▶ We drop imputed cases and individuals with self-employment income
- ▶ Final samples: 283,000 cases for DER-CPS, 161,000 for W-2-CPS

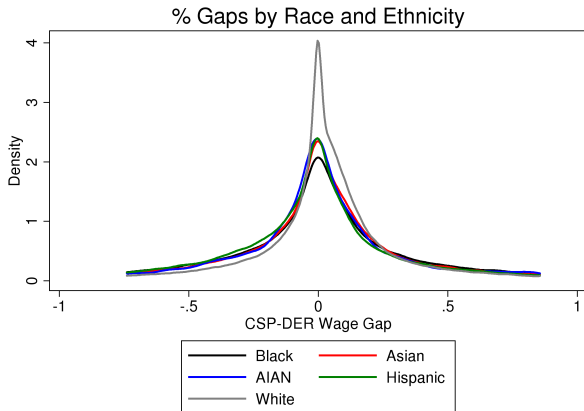
Wage Gap

- ▶ Our analysis will center on the wage reporting differential (wage gap) between survey and administrative records
- ▶ We construct the wage gap as:

$$G_{ist} = \log(Y_{ist,CPS}) - \log(Y_{ist,A})$$

- ▶ $Y_{ist,CPS}$ are wages reported to CPS
- ▶ $Y_{ist,A}$ are wages in administrative records
- ▶ Positive gap: reported more wages to CPS
- ▶ Hypotheses:
 - ▶ Do highly educated individuals misreport less?
 - ▶ Is there heterogeneity by race?

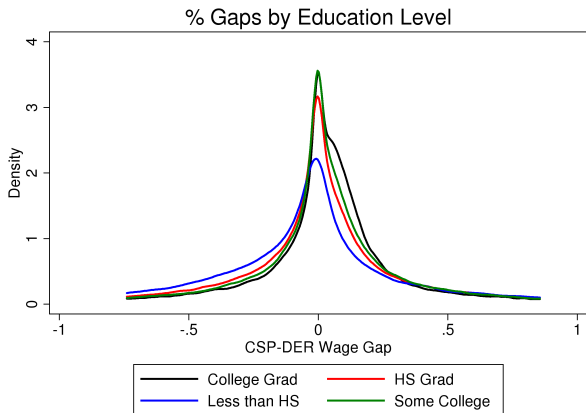
Wage Gaps by Race and Ethnicity



Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

Note: The Hispanic group includes Hispanics of any race, other race groups include only non-Hispanics

Wage Gaps by Education



Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

Wage Gap Regressions

- ▶ We further explore this heterogeneity in a regression context:
- ▶ The model:

$$G_{ist} = \alpha + \delta Demogs_{it} + \phi FE_{st} + \epsilon_{ist}$$

- ▶ $Demogs_{it}$ are socio-demographic characteristics (age, gender, marital status, race and ethnicity, education)
- ▶ FE_{st} are state and year fixed effects

W-2 and DER Wage Gap Regressions

Wage Gap:	(1) DER	(2) W-2	(3) DER	(4) W-2
Married	0.0042 (0.0061)	0.0050 (0.0079)	-0.0005 (0.0028)	-0.0000 (0.0033)
Female	0.0191** (0.0079)	0.0279** (0.0116)	0.0029 (0.0035)	0.0068 (0.0047)
Black	0.0048 (0.0143)	-0.0013 (0.0209)	-0.0123** (0.0052)	-0.0143** (0.0072)
Asian	-0.0110 (0.0240)	0.0115 (0.0276)	-0.0282** (0.0113)	-0.0118 (0.0137)
AIAN	0.0062 (0.0398)	0.0456 (0.0579)	0.0065 (0.0218)	0.0288 (0.0254)
Hispanic	-0.0504*** (0.0118)	-0.0459*** (0.0155)	-0.0353*** (0.0060)	-0.0386*** (0.0085)
Less Than High School	-0.0212 (0.0149)	-0.0162 (0.0186)	-0.0044 (0.0064)	-0.0024 (0.0087)
Some College	0.0128 (0.0095)	0.0215* (0.0115)	0.0137*** (0.0039)	0.0187*** (0.0050)
Bachelor's Degree	0.0165** (0.0073)	0.0212** (0.0098)	0.0171*** (0.0035)	0.0243*** (0.0047)
Age	-0.0027*** (0.0003)	-0.0028*** (0.0004)	-0.0015*** (0.0001)	-0.0017*** (0.0002)
Observations	283,000	161,000	254,000	145,000
State Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Percentile Range	0-100	0-100	5-95	5-95

Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

W-2 and DER Wage Gap Regressions

Wage Gap:	(1) DER	(2) W-2	(3) DER	(4) W-2
Married	0.0042 (0.0061)	0.0050 (0.0079)	-0.0005 (0.0028)	-0.0000 (0.0033)
Female	0.0191** (0.0079)	0.0279** (0.0116)	0.0029 (0.0035)	0.0068 (0.0047)
Black	0.0048 (0.0143)	-0.0013 (0.0209)	-0.0123** (0.0052)	-0.0143** (0.0072)
Asian	-0.0110 (0.0240)	0.0115 (0.0276)	-0.0282** (0.0113)	-0.0118 (0.0137)
AIAN	0.0062 (0.0398)	0.0456 (0.0579)	0.0065 (0.0218)	0.0288 (0.0254)
Hispanic	-0.0504*** (0.0118)	-0.0459*** (0.0155)	-0.0353*** (0.0060)	-0.0386*** (0.0085)
Less Than High School	-0.0212 (0.0149)	-0.0162 (0.0186)	-0.0044 (0.0064)	-0.0024 (0.0087)
Some College	0.0128 (0.0095)	0.0215* (0.0115)	0.0137*** (0.0039)	0.0187*** (0.0050)
Bachelor's Degree	0.0165** (0.0073)	0.0212** (0.0098)	0.0171*** (0.0035)	0.0243*** (0.0047)
Age	-0.0027*** (0.0003)	-0.0028*** (0.0004)	-0.0015*** (0.0001)	-0.0017*** (0.0002)
Observations	283,000	161,000	254,000	145,000
State Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Percentile Range	0-100	0-100	5-95	5-95

Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

W-2 and DER Wage Gap Regressions

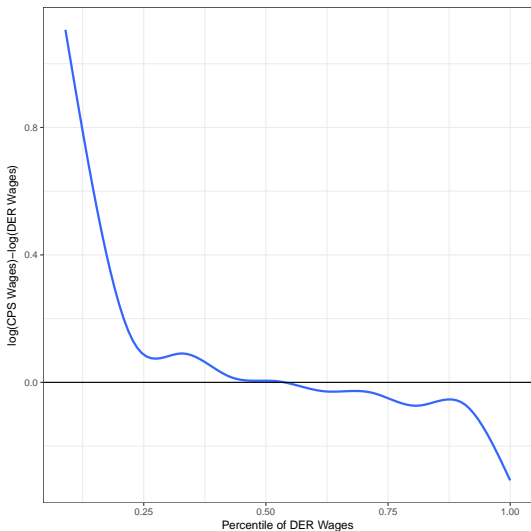
Wage Gap:	(1) DER	(2) W-2	(3) DER	(4) W-2
Married	0.0042 (0.0061)	0.0050 (0.0079)	-0.0005 (0.0028)	-0.0000 (0.0033)
Female	0.0191** (0.0079)	0.0279** (0.0116)	0.0029 (0.0035)	0.0068 (0.0047)
Black	0.0048 (0.0143)	-0.0013 (0.0209)	-0.0123** (0.0052)	-0.0143** (0.0072)
Asian	-0.0110 (0.0240)	0.0115 (0.0276)	-0.0282** (0.0113)	-0.0118 (0.0137)
AIAN	0.0062 (0.0398)	0.0456 (0.0579)	0.0065 (0.0218)	0.0288 (0.0254)
Hispanic	-0.0504*** (0.0118)	-0.0459*** (0.0155)	-0.0353*** (0.0060)	-0.0386*** (0.0085)
Less Than High School	-0.0212 (0.0149)	-0.0162 (0.0186)	-0.0044 (0.0064)	-0.0024 (0.0087)
Some College	0.0128 (0.0095)	0.0215* (0.0115)	0.0137*** (0.0039)	0.0187*** (0.0050)
Bachelor's Degree	0.0165** (0.0073)	0.0212** (0.0098)	0.0171*** (0.0035)	0.0243*** (0.0047)
Age	-0.0027*** (0.0003)	-0.0028*** (0.0004)	-0.0015*** (0.0001)	-0.0017*** (0.0002)
Observations	283,000	161,000	254,000	145,000
State Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Percentile Range	0-100	0-100	5-95	5-95

Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

Heterogeneity Across the Wage Distribution

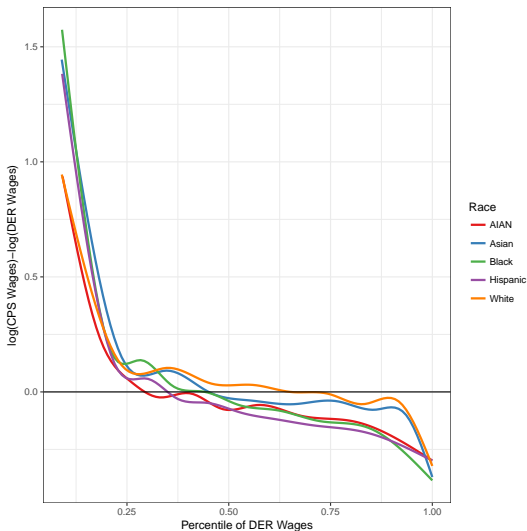
- ▶ Mis-reporting may systematically differ across the wage distribution alongside heterogeneity across groups
- ▶ To examine this, estimate average wage gaps by percentile of the DER wage distribution
 - ▶ We visualize this by fitting a bivariate Generalized Additive Model to the wage gap and wage percentile data

Heterogeneity Across the Wage Distribution



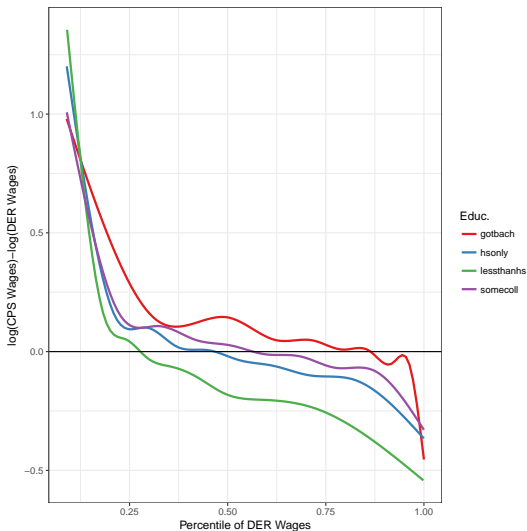
Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

Heterogeneity Across the Wage Distribution, by Race



Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

Heterogeneity Across the Wage Distribution, by Education



Source: CPS ASEC, IRS 1040, SSA DER, IRS W-2 2000-2015

Conclusion

- ▶ We provide evidence of systematic variation in misreporting across several Socio-demographic dimensions
 - ▶ Both on average and across the wage distribution
 - ▶ Educational attainment seems particularly important
- ▶ A note:
 - ▶ “wages” have become a fuzzy concept as independent contracting has increased
 - ▶ Follow-up work: looking at individuals or tax units with self-employment income

Conclusion

Thanks!

Christian Imboden
email: cimboden@uoregon.edu

John Voorheis
email: john.l.voorheis@census.gov

Caroline Weber
email: cweber5@uoregon.edu