

Reconciling Survey and Administrative Measures of Self-Employment

Katharine G. Abraham, University of Maryland

John C. Haltiwanger, University of Maryland

Claire Hou, University of Maryland & U.S. Census Bureau

Kristin Sandusky, U.S. Census Bureau

James R. Spletzer, U.S. Census Bureau

LERA session “Short-Time Work Arrangements”

January 4, 2020

Disclaimers

Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

All results have been reviewed to ensure that no confidential information is disclosed (DRB-B0046-CED-20190425 and CBDRB-FY20-CED006-0008).

Motivation: Self-employment levels and trends differ across HH surveys and admin (tax) data

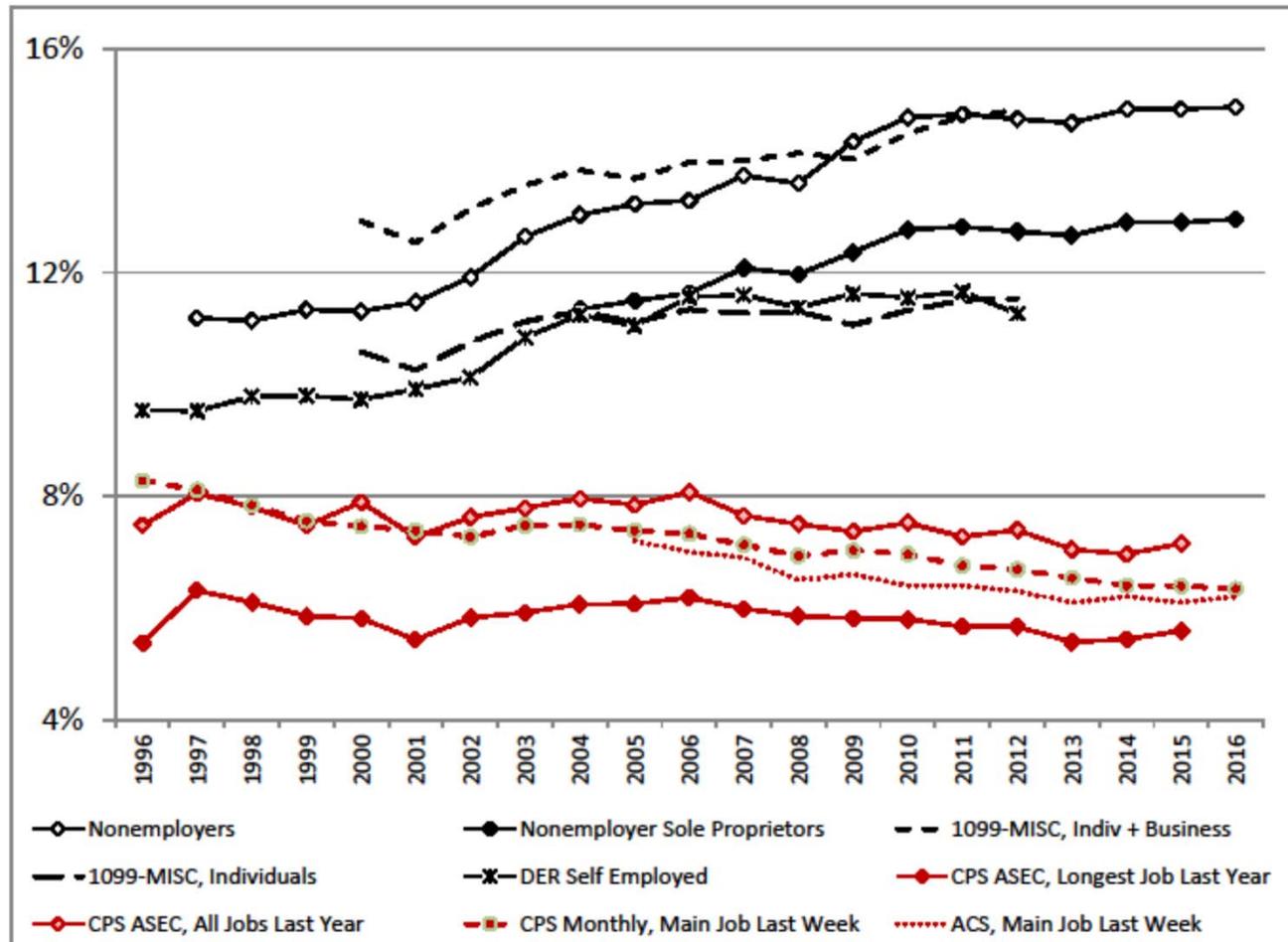


Figure 1 from AHSS 2018 (NBER WP 24950),

Why we care about accurate measurement of self-employment

Economic Measurement

- Perception of a growing “gig economy”
- Help understand the long-run decline in the employment-population ratio
- Unreported earnings and implications for GDP
- Productivity measurement

Policy Implications

- Health insurance, retirement savings, legal rights and protections
- Understanding how families make ends meet

Concern about different levels and trends of self-employment

Household surveys are invaluable

- CPS is available 2-3 weeks after fielding → invaluable for understanding aggregate labor market conditions
- Microdata are publicly available → invaluable for research

Concern about how well household surveys are capturing self-employment and other forms of non-traditional work

- Reflection of how difficult it can be to capture complex arrangements by asking a small number of questions on a household survey (*Monthly Labor Review* September 2018, Abraham & Amaya *Journal of Official Statistics* 2019)

Analyzing discrepancies in self-employment levels and trends

The best way to understand discrepancies between household survey data and administrative tax data is to compare information from the two sources for the same set of people

Our previous work (2013 JoLE) analyzed discrepancies in employment status between CPS and employer-reported administrative records:

- persons employed in the admin data but not in the CPS have personal characteristics consistent with not viewing employment as their main activity (students, retirees) and job characteristics consistent with marginal work (short term, low earnings)
- persons employed as wage and salary workers in the CPS but not in the admin data are “off-the-books” or misclassified independent contractors

In this research, we use the AHSS (2013 JoLE) methodology to study the growing discrepancies between self-employment status in the CPS-ASEC and the administrative tax data

Important caveat

Can we say that household survey data or administrative tax data are more accurate? No

- CPS & ACS don't probe deeply about non-traditional work arrangements, leading to reporting errors
- Tax data only captures what is reported to the tax authorities

To the extent there are measurement issues, our hope is that understanding the discrepancies will lead to improvements in the data and to a better understanding of the self-employed

20 years of linked CPS – admin tax data 1996-2015

CPS information from Annual Social and Economic Supplement (ASEC)

Self-employed are those with positive net self-employment income

CPS-ASEC self-employment question: What were (name's/your) net earnings from the business/farm after expenses during YEAR?

Tax information from Detailed Earnings Record (DER) provided by the Social Security Administration to the U.S. Census Bureau

DER includes information from all W-2s and Schedule SE's

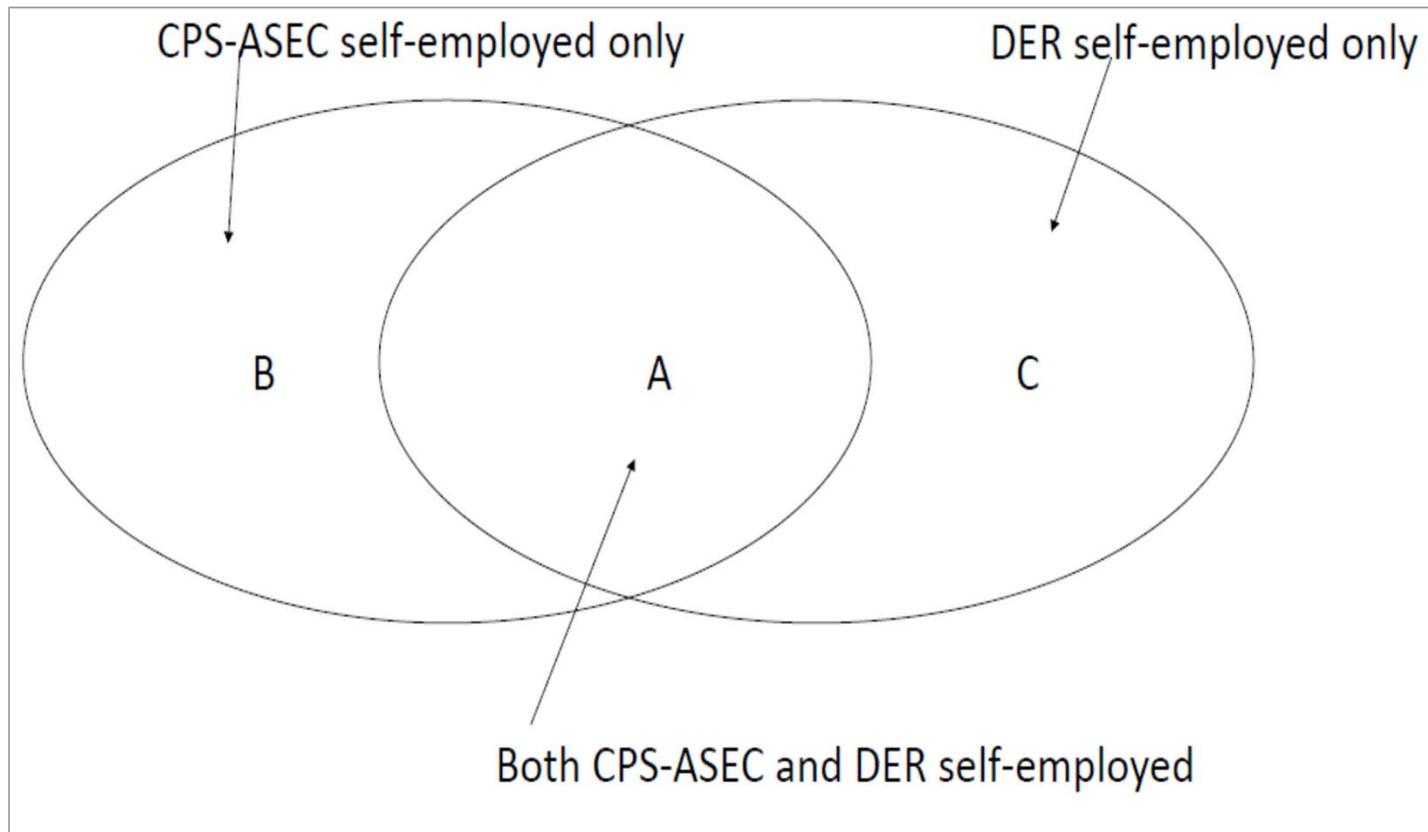
Schedule SE should be filed if \$433 or more in net self-employment earnings

Definition of self-employment should be conceptually comparable in both data sources

Both have an annual reference frame

One difference is the \$433 threshold in the tax data (\$0 in CPS-ASEC)

Self-employment in CPS-ASEC and DER



Self-employment in CPS-ASEC and DER

Annual Average, 1996-2015

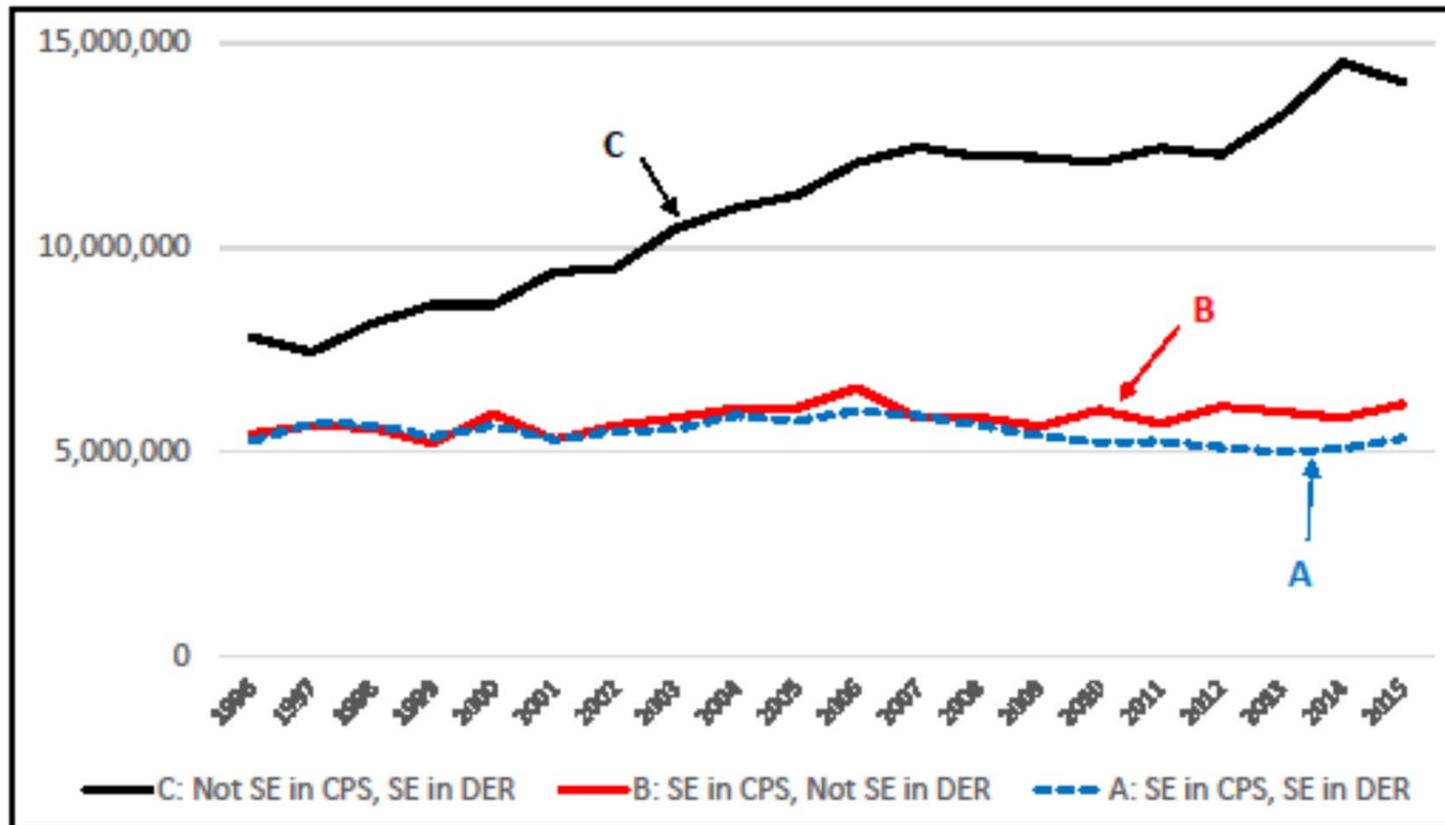
	Not self-employed in DER	Self-employed in DER	Total
<u>Not self-employed in CPS</u>			
Number	205,849,371	10,978,424	216,827,794
Row share	94.9%	5.1%	100.0%
Column share	97.3%	66.7%	95.1%
<u>Self-employed in CPS</u>			
Number	5,808,202	5,471,298	11,279,501
Row share	51.5%	48.5%	100.0%
Column share	2.7%	33.3%	4.9%
<u>Total</u>			
Number	211,657,573	16,449,722	228,107,295
Row share	92.8%	7.2%	100.0%
Column share	100.0%	100.0%	100.0%

Self-employment in CPS-ASEC and DER

Annual Average, 1996-2015

	Not self-employed in DER	Self-employed in DER	Total
<u>Not self-employed in CPS</u>			
Number	205,849,371	10,978,424	216,827,794
Row share	94.9%	5.1%	100.0%
Column share	97.3%	66.7%	95.1%
<u>Self-employed in CPS</u>			
Number	5,808,202	5,471,298	11,279,501
Row share	51.5%	48.5%	100.0%
Column share	2.7%	33.3%	4.9%
<u>Total</u>			
Number	211,657,573	16,449,722	228,107,295
Row share	92.8%	7.2%	100.0%
Column share	100.0%	100.0%	100.0%

Self-employment in CPS-ASEC and DER 1996-2015

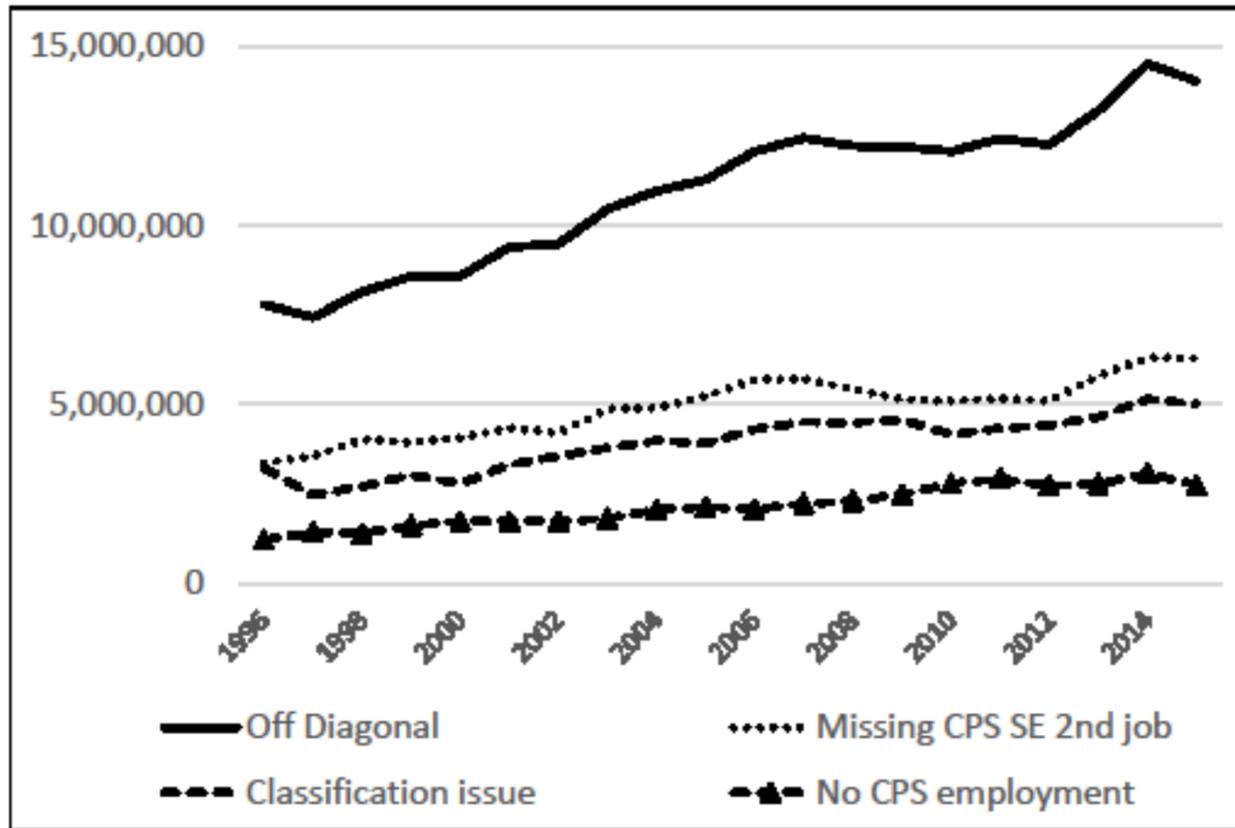


Disaggregation of “C” off-diagonal Self-employed in DER, not in CPS-ASEC

Three distinct groups of interest

- Missing CPS SE 2nd job: W&S income in both DER and CPS-ASEC, self-employment income only in DER
- Classification issue: Only self-employment income in DER, only W&S income in CPS-ASEC
- No CPS employment: Self-employment income in DER (either alone or together with W&S income), no employment income in CPS-ASEC

Self-employed in DER, not in CPS-ASEC 1996-2015



Each of the three distinct groups explains part of the growth of the off-diagonal

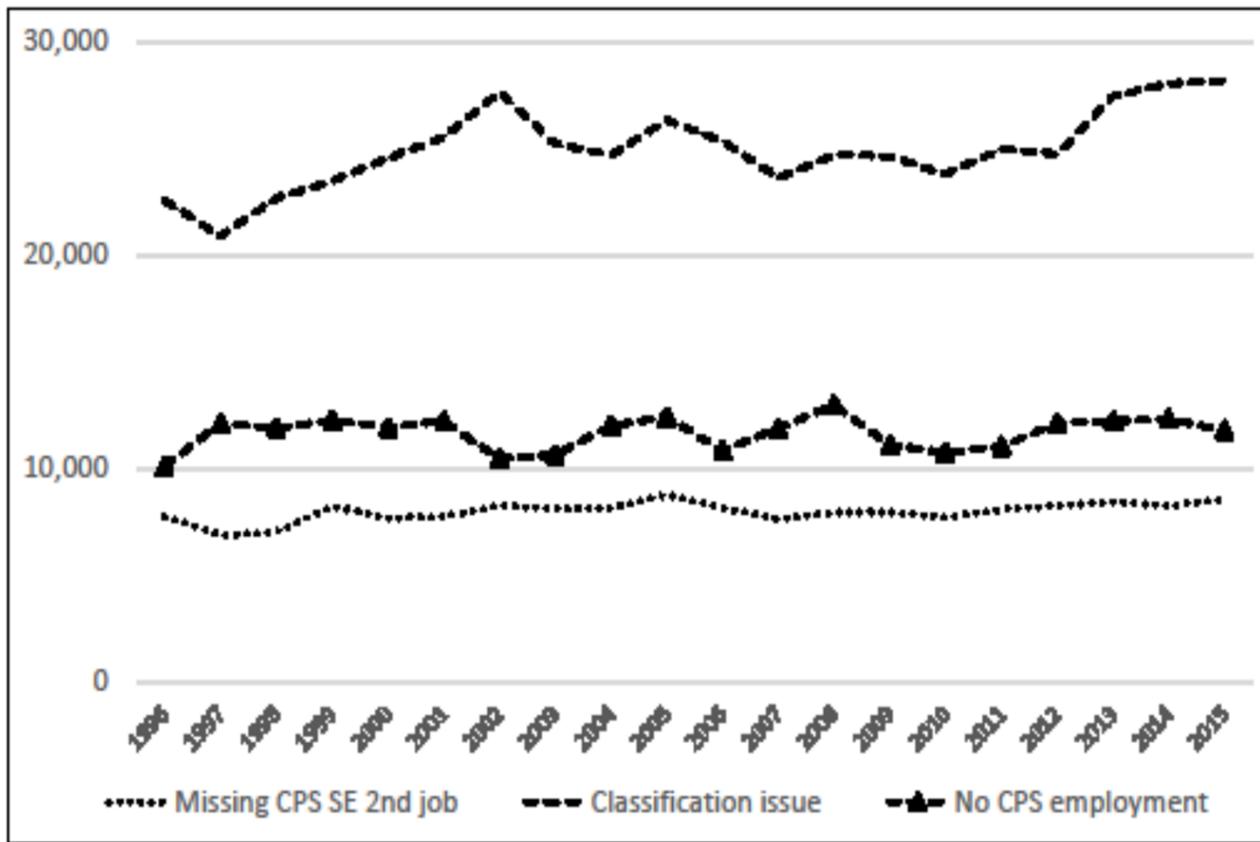
Levels:

- 45% Missing 2nd job
- 36% Classification
- 20% No CPS emp

Growth:

- 43% Missing 2nd job
- 34% Classification
- 24% No CPS emp

Real DER Self-Employment Earnings 1996-2015



Real earnings of
“misclassified”
are approx
\$25,000 per year

Real earnings of
missing CPS SE
jobs are approx
\$10,000 per year

- Note the non-decreasing trend

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Descriptive Regressions

- Sample: linked CPS-DER with positive self-employment earnings in the DER, 1996-2015
- Dependent variables: =1 if in specified category, =0 else
 - 1) No CPS self-employment (the “off-diagonal”)
 - 2) Missing CPS SE 2nd job
 - 3) Classification issue
 - 4) No CPS employment
- Explanatory variables: CPS demographics, DER self-employment earnings quartiles, year dummies, ...

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Select Coefficients	CPS SE=0	Missing SE	Misclassify	No CPS
	DER SE=1	2nd job DER SE=1		Emp DER SE=1
Age 15-24	.112 *	.065 *	.012 *	.035 *
Age 25-34	.043 *	.025 *	.005	.013 *
Age 45-54	-.032 *	-.027 *	-.004	-.002 *
Age 55-64	-.059 *	-.068 *	.005	.004 *
Age 65+	-.050 *	-.149 *	.041 *	.058 *
Enrolled in school	.030 *	-.134 *	.033 *	.130 *
NILF - retired	.263 *	-.239 *	-.104 *	.606 *
Black	.102 *	.041 *	.003	.058 *
Other race	.010 *	-.017 *	.010 *	.017 *
Foreign Born	.084 *	-.027 *	.110 *	.002 *
Male	-.002	.036 *	.012 *	-.050 *
Proxy response	.054 *	.021 *	.016 *	.018 *
DER SE \$ Q2	-.067 *	-.108 *	.039 *	.002
DER SE \$ Q3	-.121 *	-.245 *	.106 *	.017 *
DER SE \$ Q4	-.173 *	-.329 *	.158 *	-.002 *
Curr year only SE	.180 *	.233 *	-.082 *	.028 *
R-squared	.122	.204	.058	.254
Mean Dep Var	.668	.305	.232	.115

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Select Coefficients	CPS SE=0		Missing SE 2nd job		Misclassify		No CPS Emp	
	DER SE=1		DER SE=1		DER SE=1		DER SE=1	
Age 15-24	.112 *		.065 *		.012 *		.035 *	
Age 25-34	.043 *		.025 *		.005		.013 *	
Age 45-54	-.032 *		-.027 *		-.004		-.002 *	
Age 55-64	-.059 *		-.068 *		.005		.004 *	
Age 65+	-.050 *		-.149 *		.041 *		.058 *	
Enrolled in school	.030 *		-.134 *		.033 *		.130 *	
NILF - retired	.263 *		-.239 *		-.104 *		.606 *	
Black	.102 *		.041 *		.003		.058 *	
Other race	.010 *		-.017 *		.010 *		.017 *	
Foreign Born	.084 *		-.027 *		.110 *		.002 *	
Male	-.002		.036 *		.012 *		-.050 *	
Proxy response	.054 *		.021 *		.016 *		.018 *	
DER SE \$ Q2	-.067 *		-.108 *		.039 *		.002	
DER SE \$ Q3	-.121 *		-.245 *		.106 *		.017 *	
DER SE \$ Q4	-.173 *		-.329 *		.158 *		-.002 *	
Curr year only SE	.180 *		.233 *		-.082 *		.028 *	
R-squared	.122		.204		.058		.254	
Mean Dep Var	.668		.305		.232		.115	

Descriptive Regressions

DER SE who are enrolled in school are 13 pp more likely to report no CPS employment

DER SE who are retired are substantially more likely to report no CPS employment

This is consistent with our hypothesis that SE is not the main activity of these persons

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Select Coefficients	CPS SE=0		Missing SE		Misclassify		No CPS Emp	
	DER SE=1		DER SE=1		DER SE=1		DER SE=1	
Age 15-24	.112	*	.065	*	.012	*	.035	*
Age 25-34	.043	*	.025	*	.005		.013	*
Age 45-54	-.032	*	-.027	*	-.004		-.002	*
Age 55-64	-.059	*	-.068	*	.005		.004	*
Age 65+	-.050	*	-.149	*	.041	*	.058	*
Enrolled in school	.030	*	-.134	*	.033	*	.130	*
NILF - retired	.263	*	-.239	*	-.104	*	.606	*
Black	.102	*	.041	*	.003		.058	*
Other race	.010	*	-.017	*	.010	*	.017	*
Foreign Born	.084	*	-.027	*	.110	*	.002	*
Male	-.002		.036	*	.012	*	-.050	*
Proxy response	.054	*	.021	*	.016	*	.018	*
DER SE \$ Q2	-.067	*	-.108	*	.039	*	.002	
DER SE \$ Q3	-.121	*	-.245	*	.106	*	.017	*
DER SE \$ Q4	-.173	*	-.329	*	.158	*	-.002	*
Curr year only SE	.180	*	.233	*	-.082	*	.028	*
R-squared	.122		.204		.058		.254	
Mean Dep Var	.668		.305		.232		.115	

Descriptive Regressions

Foreign born persons more likely to report (CPS SE=0, DER SE=1)

- This effect entirely in “misclassification”

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Select Coefficients	CPS SE=0		Missing SE		Misclassify		No CPS Emp	
	DER SE=1		DER SE=1		DER SE=1		DER SE=1	
Age 15-24	.112	*	.065	*	.012	*	.035	*
Age 25-34	.043	*	.025	*	.005		.013	*
Age 45-54	-.032	*	-.027	*	-.004		-.002	*
Age 55-64	-.059	*	-.068	*	.005		.004	*
Age 65+	-.050	*	-.149	*	.041	*	.058	*
Enrolled in school	.030	*	-.134	*	.033	*	.130	*
NILF - retired	.263	*	-.239	*	-.104	*	.606	*
Black	.102	*	.041	*	.003		.058	*
Other race	.010	*	-.017	*	.010	*	.017	*
Foreign Born	.084	*	-.027	*	.110	*	.002	*
Male	-.002		.036	*	.012	*	-.050	*
Proxy response	.054	*	.021	*	.016	*	.018	*
DER SE \$ Q2	-.067	*	-.108	*	.039	*	.002	
DER SE \$ Q3	-.121	*	-.245	*	.106	*	.017	*
DER SE \$ Q4	-.173	*	-.329	*	.158	*	-.002	*
Curr year only SE	.180	*	.233	*	-.082	*	.028	*
R-squared	.122		.204		.058		.254	
Mean Dep Var	.668		.305		.232		.115	

Descriptive Regressions

CPS Proxy respondents are more likely to report (CPS SE=0, DER SE=1)

- But effect is arguably small (5.4 pp of the 66.8% mean)

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Select Coefficients	CPS SE=0		Missing SE 2nd job		Misclassify		No CPS Emp	
	DER SE=1		DER SE=1		DER SE=1		DER SE=1	
Age 15-24	.112 *		.065 *		.012 *		.035 *	
Age 25-34	.043 *		.025 *		.005		.013 *	
Age 45-54	-.032 *		-.027 *		-.004		-.002 *	
Age 55-64	-.059 *		-.068 *		.005		.004 *	
Age 65+	-.050 *		-.149 *		.041 *		.058 *	
Enrolled in school	.030 *		-.134 *		.033 *		.130 *	
NILF - retired	.263 *		-.239 *		-.104 *		.606 *	
Black	.102 *		.041 *		.003		.058 *	
Other race	.010 *		-.017 *		.010 *		.017 *	
Foreign Born	.084 *		-.027 *		.110 *		.002 *	
Male	-.002		.036 *		.012 *		-.050 *	
Proxy response	.054 *		.021 *		.016 *		.018 *	
DER SE \$ Q2	-.067 *		-.108 *		.039 *		.002	
DER SE \$ Q3	-.121 *		-.245 *		.106 *		.017 *	
DER SE \$ Q4	-.173 *		-.329 *		.158 *		-.002 *	
Curr year only SE	.180 *		.233 *		-.082 *		.028 *	
R-squared	.122		.204		.058		.254	
Mean Dep Var	.668		.305		.232		.115	

Descriptive Regressions

DER SE are more likely to report their 2nd SE job to the CPS when it is high earnings

- Consistent with marginal work

DER SE with high earnings are more likely to be in the misclassified group

Who is reporting self-employment income in DER but not in the CPS-ASEC?

Select Coefficients	CPS SE=0	Missing SE	Misclassify	No CPS
	DER SE=1	2nd job DER SE=1		Emp DER SE=1
Age 15-24	.112 *	.065 *	.012 *	.035 *
Age 25-34	.043 *	.025 *	.005	.013 *
Age 45-54	-.032 *	-.027 *	-.004	-.002 *
Age 55-64	-.059 *	-.068 *	.005	.004 *
Age 65+	-.050 *	-.149 *	.041 *	.058 *
Enrolled in school	.030 *	-.134 *	.033 *	.130 *
NILF - retired	.263 *	-.239 *	-.104 *	.606 *
Black	.102 *	.041 *	.003	.058 *
Other race	.010 *	-.017 *	.010 *	.017 *
Foreign Born	.084 *	-.027 *	.110 *	.002 *
Male	-.002	.036 *	.012 *	-.050 *
Proxy response	.054 *	.021 *	.016 *	.018 *
DER SE \$ Q2	-.067 *	-.108 *	.039 *	.002
DER SE \$ Q3	-.121 *	-.245 *	.106 *	.017 *
DER SE \$ Q4	-.173 *	-.329 *	.158 *	-.002 *
Curr year only SE	.180 *	.233 *	-.082 *	.028 *
R-squared	.122	.204	.058	.254
Mean Dep Var	.668	.305	.232	.115

Descriptive Regressions

“Current year only SE” is a {0,1} variable constructed from the longitudinal DER (=1 if the person is SE in only the current year but not the previous nor following years)

- Consistent with short duration SE being considered marginal work

Who is reporting self-employment income in DER but not in the CPS-ASEC?

	CPS SE=0 DER SE=1	Missing SE 2nd job DER SE=1	Misclassify DER SE=1	No CPS Emp DER SE=1
ΔY , 1996/97 - 2014/15	.151	.053	.050	.048
$\Delta X\beta/\Delta Y$ (X=demog)	12.4%	-10.7%	27.8%	21.6%
$\Delta X\beta/\Delta Y$ (X=DER SE vars)	3.6%	15.8%	-7.2%	1.7%
$\Delta X\beta/\Delta Y$ (all X \neq year)	16.1%	5.1%	20.6%	23.3%

Descriptive Regressions

We compute $\Delta X\beta/\Delta Y$, which shows that changes in the “X” variables account for 16.1% of the growth in the off-diagonal

Summary: the “C” off-diagonal

Motivating Facts:

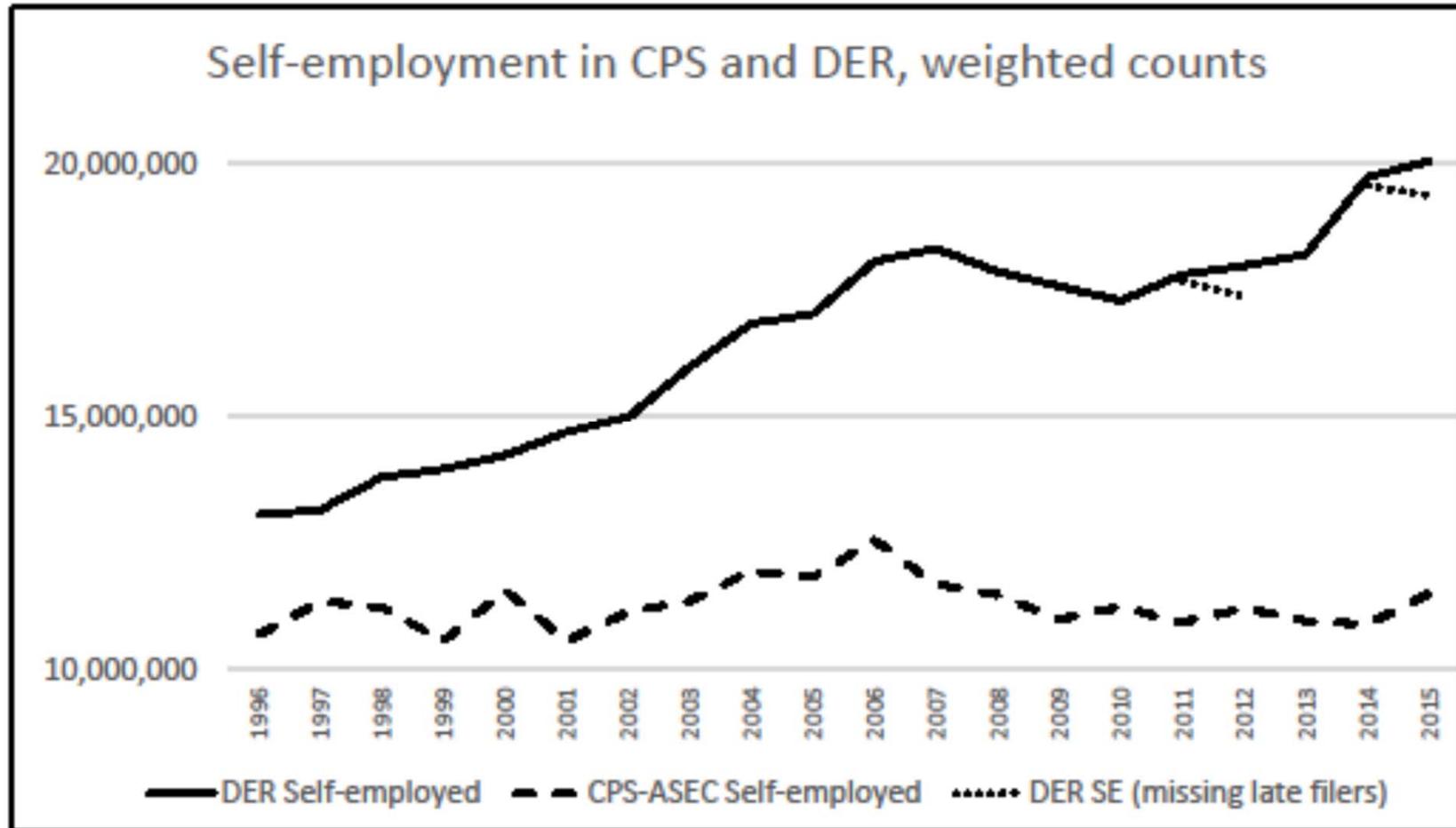
- On average over 1996-2015, two-thirds of those with self-employment income in the administrative tax data do not have self-employment income in the CPS-ASEC
- Significant growth during 1996-2015 in this off-diagonal (DER SE=1, CPS SE=0)

Using a simple descriptive regression framework:

- Persons in the “missing CPS SE 2nd job” group have characteristics consistent with self-employment being marginal work (low SE earnings, short duration SE)
- Persons in the “no CPS employment” group have characteristics consistent with something other than self-employment being their primary activity (students, retirees)
- Misclassified are different than missing: misclassified DER self-employment have high earnings and are more likely to be foreign born
- Changes in the “X” variables account for 16.1% of the growth in the off-diagonal

Extra Slides

Self-employment in CPS-ASEC and DER 1996-2015



Creating the Linked CPS-DER Data

Match CPS-ASEC and DER records by PIK

- PIK = Protected Identity Key (Census Bureau's internal individual identifier)

PIKs available for only a subset of CPS ASEC respondents (70-80% in early years, ~90% in recent years)

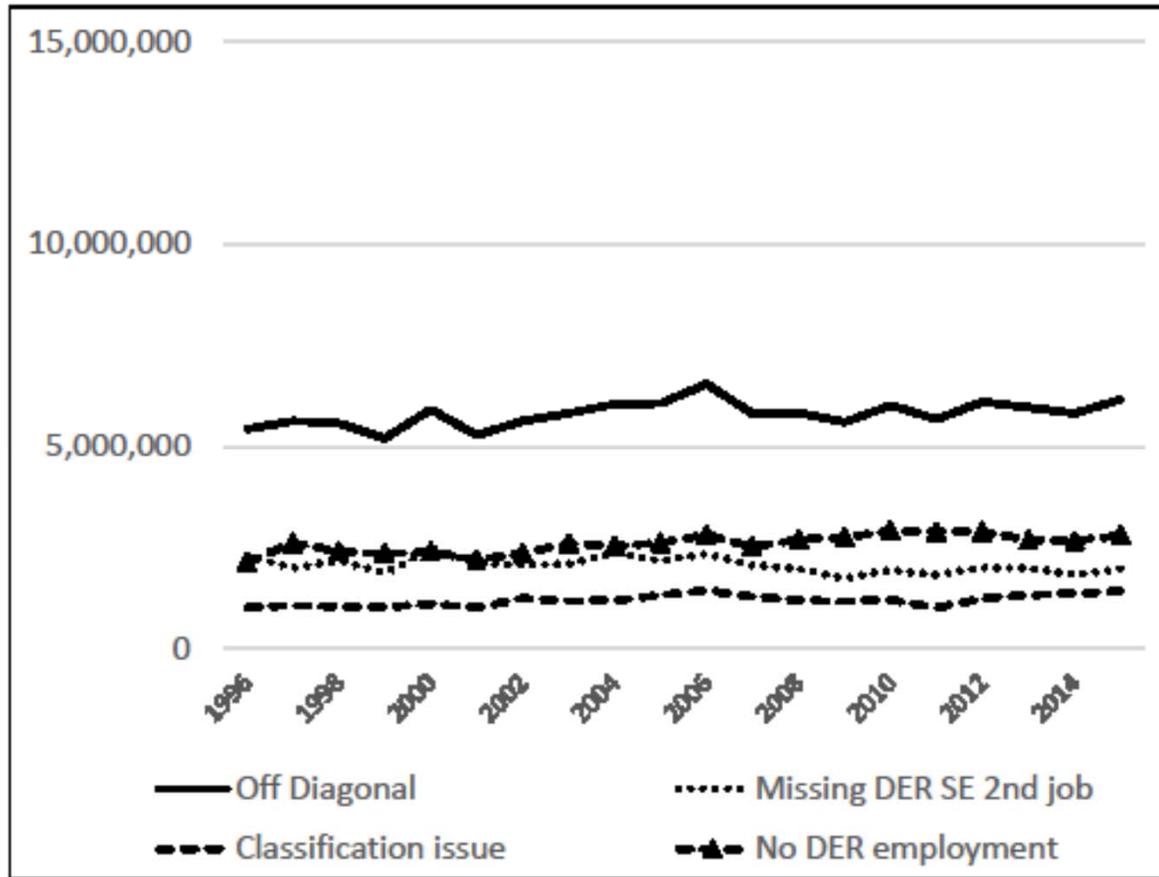
- Use propensity score methods to adjust CPS ASEC weights to account for missing PIKs

Linked sample is 65,000 - 140,000 cases per year (varies by year), weighted to represent the population as a whole

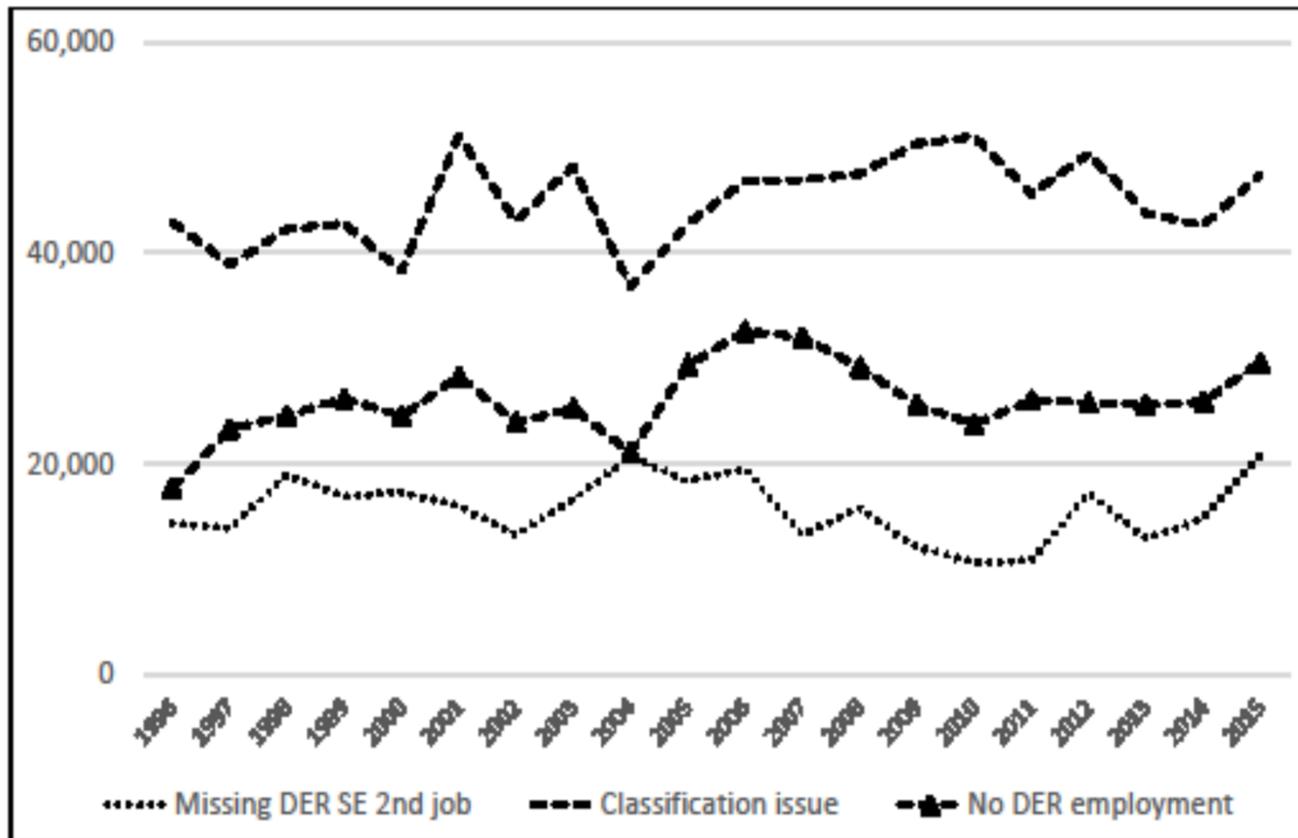
Who is reporting self-employment income in DER but not in the CPS-ASEC?

	CPS SE=0 DER SE=1	Missing SE 2nd job DER SE=1	Misclassify DER SE=1	No CPS Emp DER SE=1
ΔY , 1996/97 - 2014/15	.151	.053	.050	.048
$\Delta X\beta/\Delta Y$ (X=demog)	12.4%	-10.7%	27.8%	21.6%
$\Delta X\beta/\Delta Y$ (X=DER SE vars)	3.6%	15.8%	-7.2%	1.7%
$\Delta X\beta/\Delta Y$ (all X \neq year)	16.1%	5.1%	20.6%	23.3%
ΔY , 1996/97 - 2006/07	.092	.049	.025	.017
$\Delta X\beta/\Delta Y$ (X=demog)	15.1%	5.2%	44.4%	-2.0%
$\Delta X\beta/\Delta Y$ (X=DER SE vars)	4.8%	11.2%	-7.8%	5.0%
$\Delta X\beta/\Delta Y$ (all X \neq year)	19.9%	16.5%	36.6%	4.9%
ΔY , 2006/07 - 2009/10	.022	-.019	.008	.033
$\Delta X\beta/\Delta Y$ (X=demog)	6.3%	5.7%	-8.1%	9.4%
$\Delta X\beta/\Delta Y$ (X=DER SE vars)	1.9%	-8.4%	-15.8%	0.3%
$\Delta X\beta/\Delta Y$ (all X \neq year)	8.2%	-2.7%	-24.0%	9.7%
ΔY , 2009/10 - 2014/15	.037	.022	.017	-.002
$\Delta X\beta/\Delta Y$ (X=demog)	9.6%	-32.1%	19.6%	-362.3%
$\Delta X\beta/\Delta Y$ (X=DER SE vars)	1.8%	5.6%	-2.4%	8.0%
$\Delta X\beta/\Delta Y$ (all X \neq year)	11.4%	-26.5%	17.2%	-354.2%

The “B” off-diagonal: Self-employed in CPS-ASEC, not in DER, 1996-2015



Real CPS-ASEC Self-Employment Earnings 1996-2015



Who is reporting self-employment income in CPS-ASEC but not in the DER?

Select Coefficients	DER SE=0	Missing SE 2nd job	Misclassify	No DER Emp
Age 15-24	.157 *	.143 *	.065 *	-.050 *
Age 25-34	.026 *	.029 *	.017 *	-.020 *
Age 45-54	-.032 *	-.019 *	-.017 *	.004
Age 55-64	-.063 *	-.057 *	-.030 *	.025 *
Age 65+	-.034 *	-.127 *	-.042 *	.135 *
Enrolled in school	.034 *	-.064 *	-.026 *	.123 *
NILF - retired	.061 *	-.142 *	.046 *	.158 *
Black	.149 *	.053 *	.037 *	.059 *
Other race	.043 *	.006	.029 *	.007
Foreign Born	-.004	-.031 *	.021 *	.006
Male	.003	.047 *	-.013 *	-.031 *
Proxy response	.022 *	-.001	.024 *	.000
CPS SE \$ Q2	-.213 *	-.261 *	.051 *	-.004
CPS SE \$ Q3	-.302 *	-.354 *	.073 *	-.021 *
CPS SE \$ Q4	-.332 *	-.377 *	.096 *	-.052 *
R-squared	.111	.192	.022	.045
Mean Dep Var				

Wage & Salary in CPS-ASEC and DER Annual Average, 1996-2012

	Not W&S in DER	W&S in DER	
Not W&S in CPS			
Number	67,171,649	12,970,069	80,141,718
Row Share	83.8%	16.2%	
Column Share	79.0%	9.3%	35.6%
W&S in CPS			
Number	17,884,195	126,907,981	144,792,175
Row Share	12.4%	87.6%	
Column Share	21.0%	90.7%	64.4%
	85,055,844	139,878,049	224,078,859
	37.8%	62.2%	