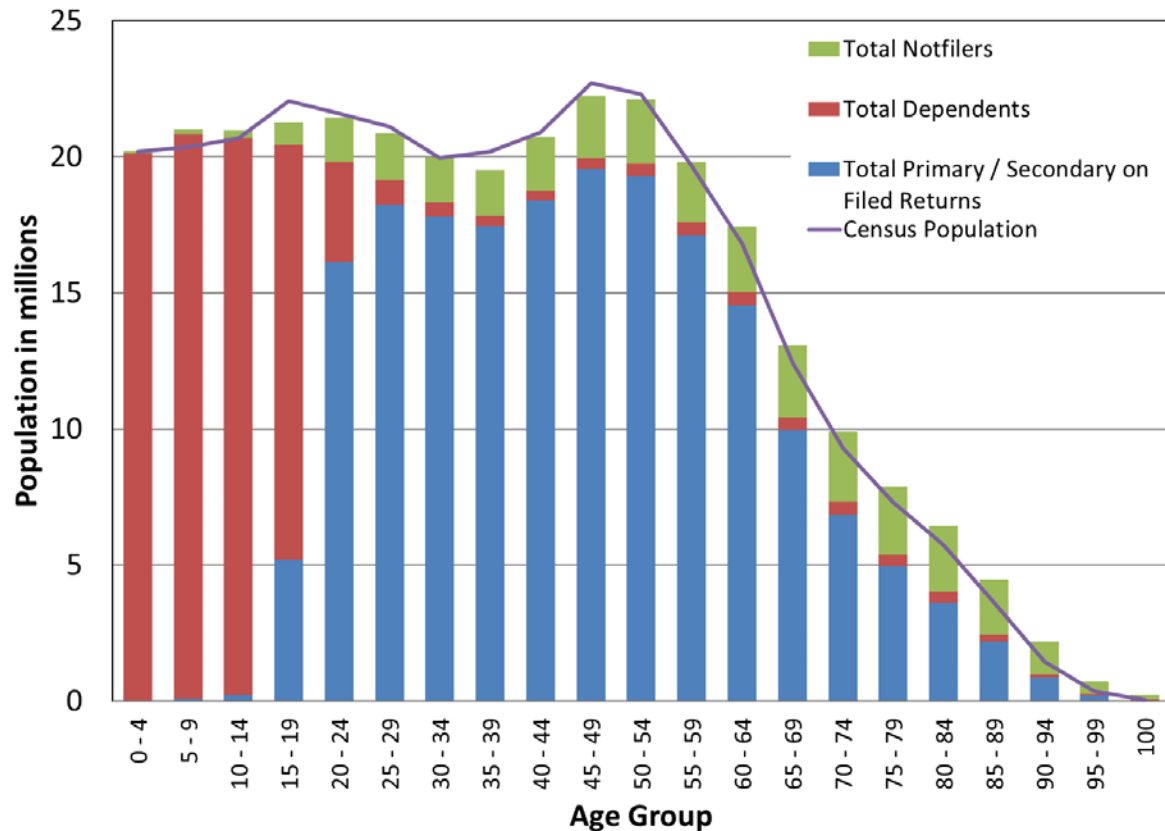# Estimating the Extent of Individual Income Tax Filing Noncompliance

Pat Langetieg, Mark Payne, and Alan Plumley
IRS Research, Applied Analytics, and Statistics: Knowledge Development & Application Division

IRS-Census Workshop on Income Measurement

March 20, 2018

IRS Administrative Population vs. Census Population by Age, 2010

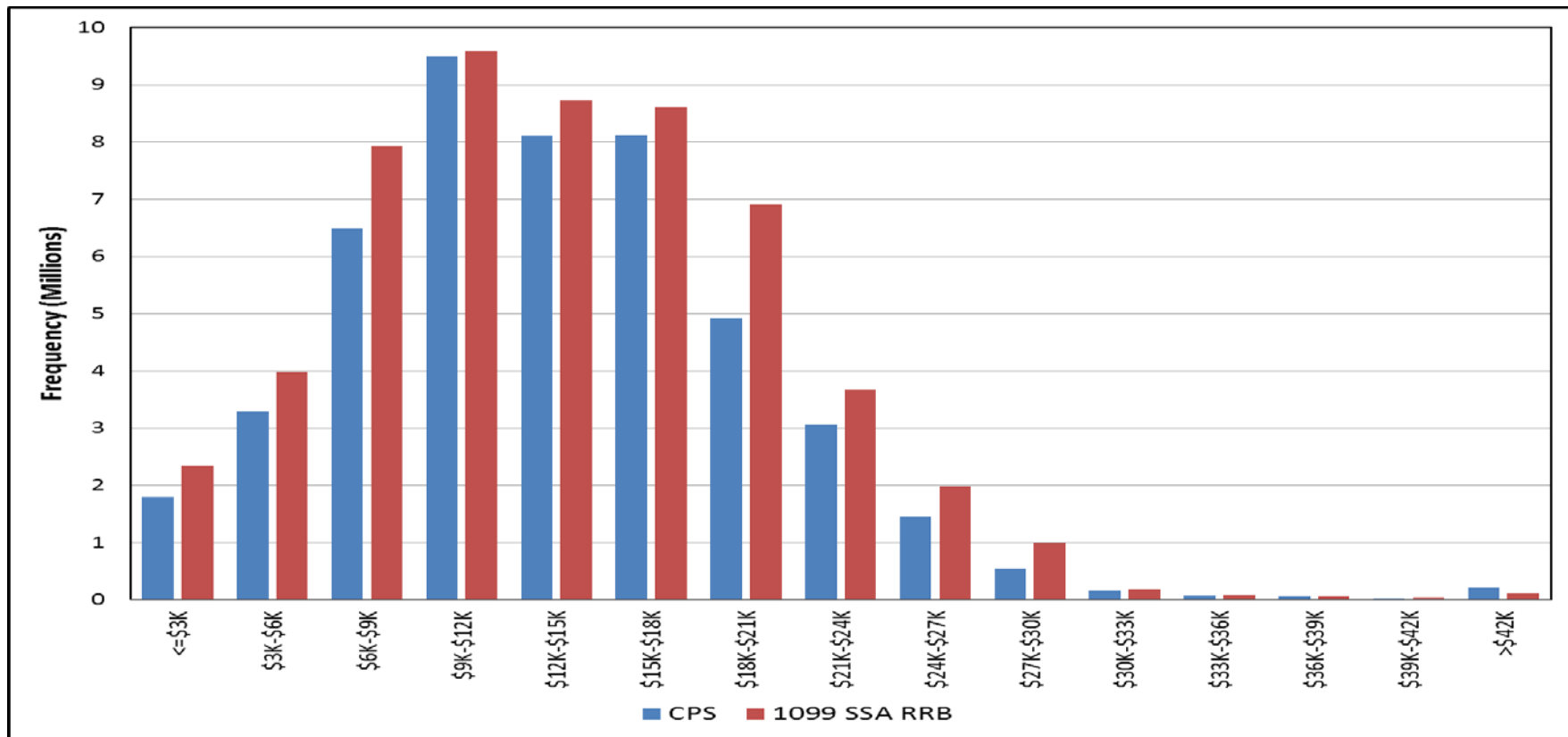**IRS population is fairly close to US Census population estimates**

IRS and Census Data Have Complementary Pros and Cons For Estimating the Extent of Nonfiling

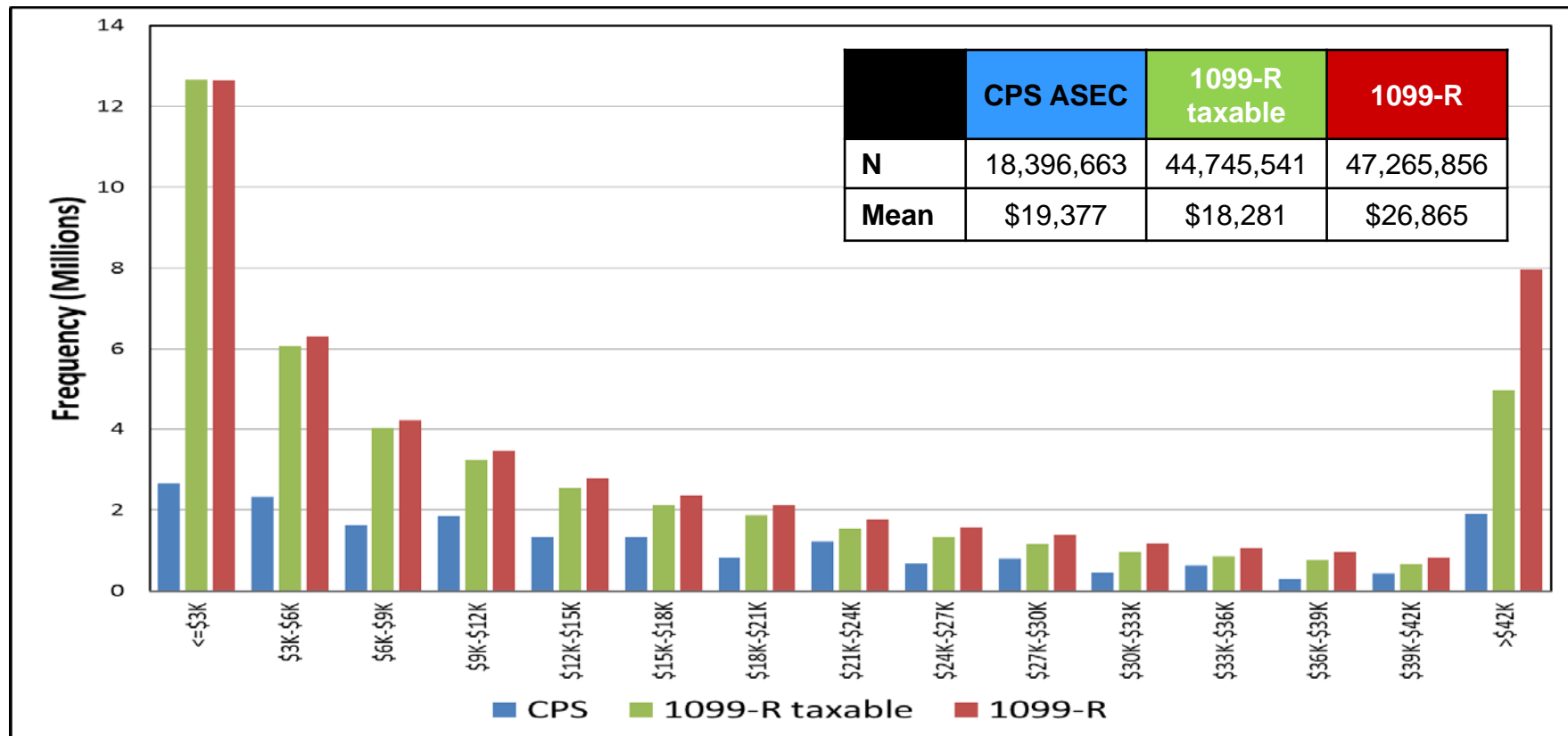|	IRS Data	Census Data
Family characteristics of nonfilers	Poor	Rich demographics
Income of nonfilers	Rich 3rd-party information returns	Poor

Therefore, we have used two approaches:

- **Administrative Data Method:** impute Census demographics to IRS data
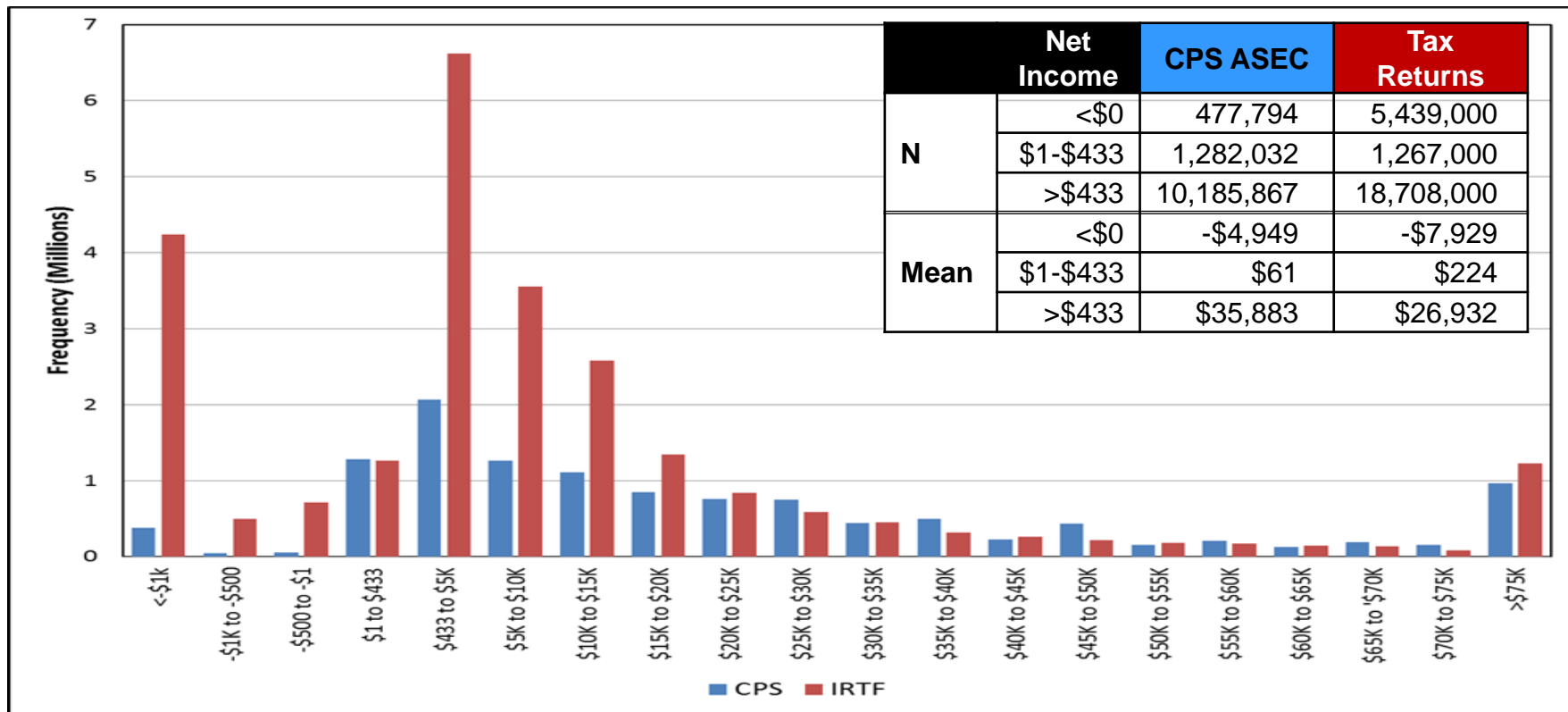- **Census Method:** impute IRS income to Census data

Research, Applied Analytics, & Statistics

Social security income not heavily underreported in the CPS

But incidence of pension income much lower in the CPS

	CPS ASEC	1099-R taxable	1099-R
N	18,396,663	44,745,541	47,265,856
Mean	$19,377	$18,281	$26,865

Self-employment income underreported even compared to tax returns

	Net Income	CPS ASEC	Tax Returns
N	<$0	477,794	5,439,000
	$1-$433	1,282,032	1,267,000
	>$433	10,185,867	18,708,000
Mean	<$0	-$4,949	-$7,929
	$1-$433	$61	$224
	>$433	$35,883	$26,932

Unemployment compensation also underreported in the CPS

	CPS ASEC	1099-G
N	8,787,926	14,990,908
Mean	$6,327	$5,655
Median	$4,200	$3,939
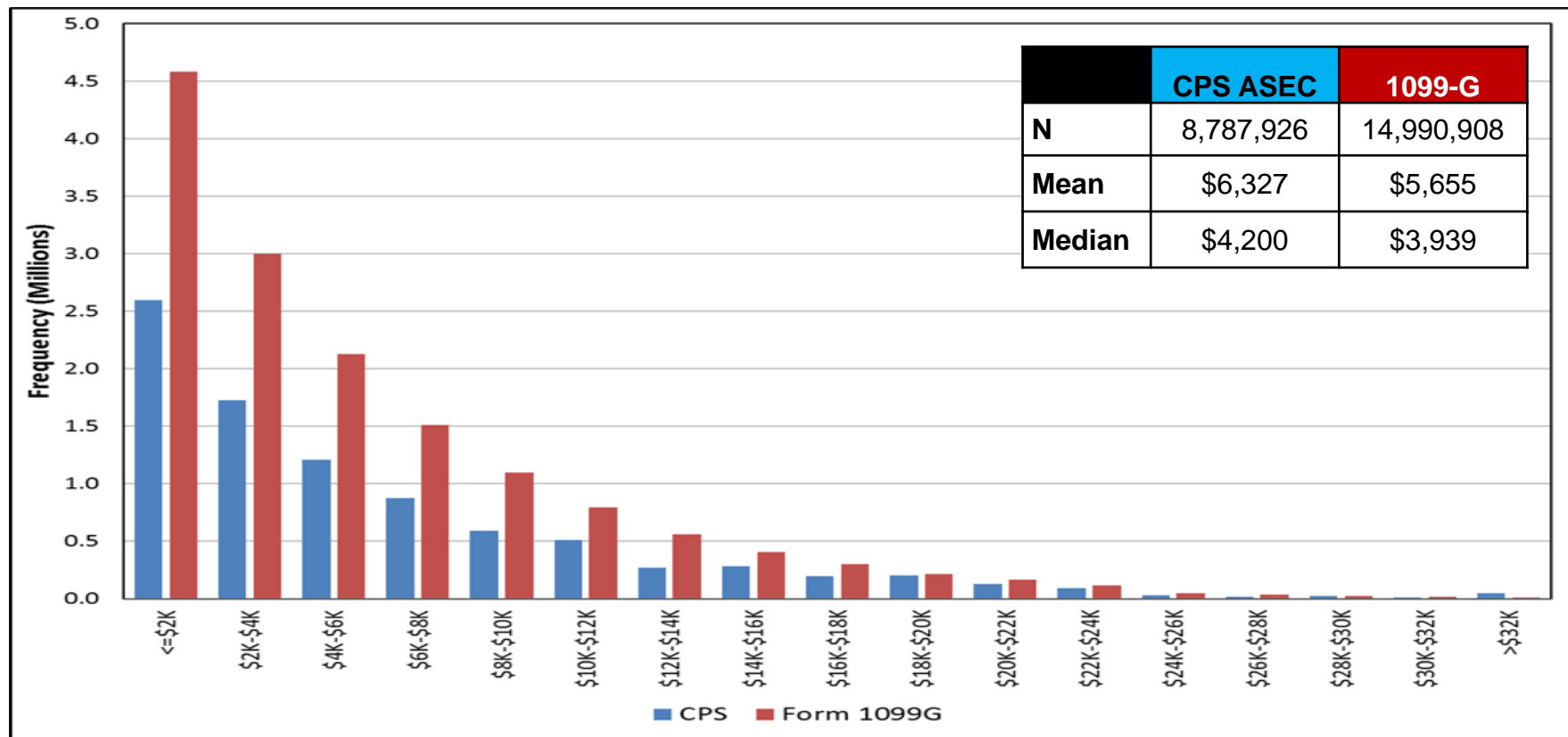
After imputations, the nonfiling tax gap estimates for the two methods are close [§]

The nonfiling tax gap is the amount of tax not paid on time by those who do not file on time. It includes both late filers and "not-filers."

Key Items	Administrative Data Method	Census Method	Amount Difference	Percentage Difference
Taxable income	$365.4	$397.9	-$32.5	-8.9%
Net tax due	$77.3	$80.8	-$3.6	-4.6%
Total payments of tax*	$48.4	$50.1	-$1.7	-3.5%
Total nonfiling gap	**$28.9**	**$30.8**	**-$1.9**	**-6.5%**

[§] Estimates averaged over Tax Years 2008 through 2010.
* Income (and tax) offsets were limited to the amount needed to reduce income (or tax) to zero.

What about the number of nonfilers?

The Voluntary Filing Rate is defined as:

$$VFR = \frac{\text{Number of Required Returns Filed on Time}}{\text{Total Number of Returns Required to be Filed}}$$

Numerator: estimated from IRS population data containing all filed returns.

Denominator:

- Census Method: estimated from the Census Bureau's *Current Population Survey, Annual Social and Economic Supplement* (CPS-ASEC).

- Administrative Data Method: estimated from IRS data

Research, Applied Analytics, & Statistics

Limitations of Census Data for Estimating Required Returns

Thousands of Returns in VFR Components Estimated by Different Methods, TY 2010

Component	Mixed Data $\dfrac{Admin}{Census}$	Single-Source Data	
		$\dfrac{Census\ Matched}{Census\ Matched}$	$\dfrac{Admin}{Admin}$
Numerator (required returns filed on time)	115,900		
Denominator (total required returns)	122,200		
Difference (implied number of nonfilers)	6,300		
Numerator/Denominator (implied VFR)	94.8%		

Limitations of Census Data for Estimating Required Returns

Thousands of Returns in VFR Components Estimated by Different Methods, TY 2010

Component	Mixed Data	Single-Source Data	
	$\dfrac{Admin}{Census}$	$\dfrac{Census\ Matched}{Census\ Matched}$	$\dfrac{Admin}{Admin}$
Numerator (required returns filed on time)	115,900	105,001	
Denominator (total required returns)	122,200	119,967	
Difference (implied number of nonfilers)	6,300	14,966	
Numerator/Denominator (implied VFR)	94.8%	87.5%	

Limitations of Census Data for Estimating Required Returns

Thousands of Returns in VFR Components Estimated by Different Methods, TY 2010

Component	Mixed Data	Single-Source Data	
	$\dfrac{Admin}{Census}$	$\dfrac{Census\ Matched}{Census\ Matched}$	$\dfrac{Admin}{Admin}$
Numerator (required returns filed on time)	115,900	105,001	115,900
Denominator (total required returns)	122,200	119,967	130,787
Difference (implied number of nonfilers)	6,300	14,966	14,937
Numerator/Denominator (implied VFR)	94.8%	87.5%	88.6%

Research, Applied Analytics, & Statistics

Limitations of Census Data for Estimating Required Returns

Thousands of Returns in VFR Components Estimated by Different Methods, TY 2010

Component	Mixed Data	Single-Source Data	
	Admin / Census	*Census Matched / Census Matched*	*Admin / Admin*
Numerator (required returns filed on time)	115,900	105,001	115,900
Denominator (total required returns)	122,200	119,967	130,787
Difference (implied number of nonfilers)	6,300	14,966	14,937
Numerator/Denominator (implied VFR)	94.8%	87.5%	88.6%

Limitations of Census Data for Estimating Required Returns

Thousands of Returns in VFR Components Estimated by Different Methods, TY 2010

Census data understate the number of required returns due to understating income.

Component	Mixed Data	Single-Source Data	
	$\dfrac{Admin}{Census}$	$\dfrac{Census\ Matched}{Census\ Matched}$	$\dfrac{Admin}{Admin}$
Numerator (required returns filed on time)	115,900	105,001	115,900
Denominator (total required returns)	122,200	119,967	130,787
Difference (implied number of nonfilers)	6,300	14,966	14,937
Numerator/Denominator (implied VFR)	94.8%	87.5%	88.6%

Efforts to Correct CPS-Based Underestimates of Required Population Have Failed

- Base total income on the 1040 amount when available; OR

- Backend imputation of gross income calibrated to totals in IRS data

Neither method is satisfactory.
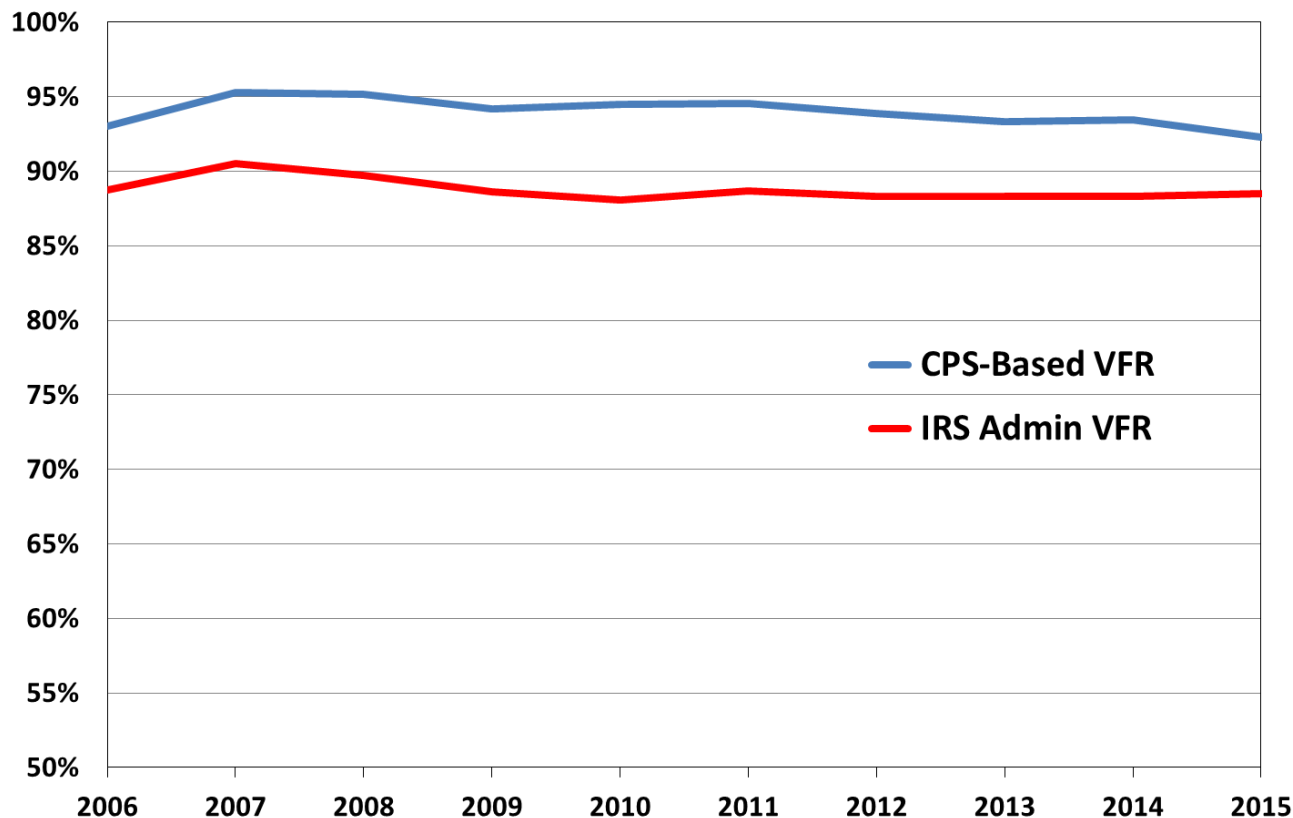
IRS Administrative Method

- **Timely and late required filers:**
  - Determine whether timely or late and whether required or not based on gross income and net self employment thresholds.
  - For consistent series, taxpayers filing more than two years after the end of tax year are treated as not-filers.

- **Not-filers (all others):**
  - On information return but not on tax return (by two year cutoff)
  - Impute net self-employment income (based on $ reported among filers).
  - Gross up net self employment income < $433.
  - Randomly assign individuals to tax units based on CPS.
  - Determine whether required to file – same as timely and late filers.

Research, Applied Analytics, & Statistics

IRS Administrative vs. CPS-Based VFR, TY 2006 to TY 2015

6.5% (~8 million) larger required population, results in VFR estimate that is about 5% lower than CPS method

Note: The 2015 IRS Administrative Data-based VFR is provisional.

Research, Applied Analytics, & Statistics

Characteristics of Nonfilers and Drivers of Nonfiling

- Since it uses the same data source for the numerator and denominator, the IRS administrative data method facilitates examination of the causes of VFR fluctuations.

- In addition, this method can facilitate learning about drivers of nonfiling.
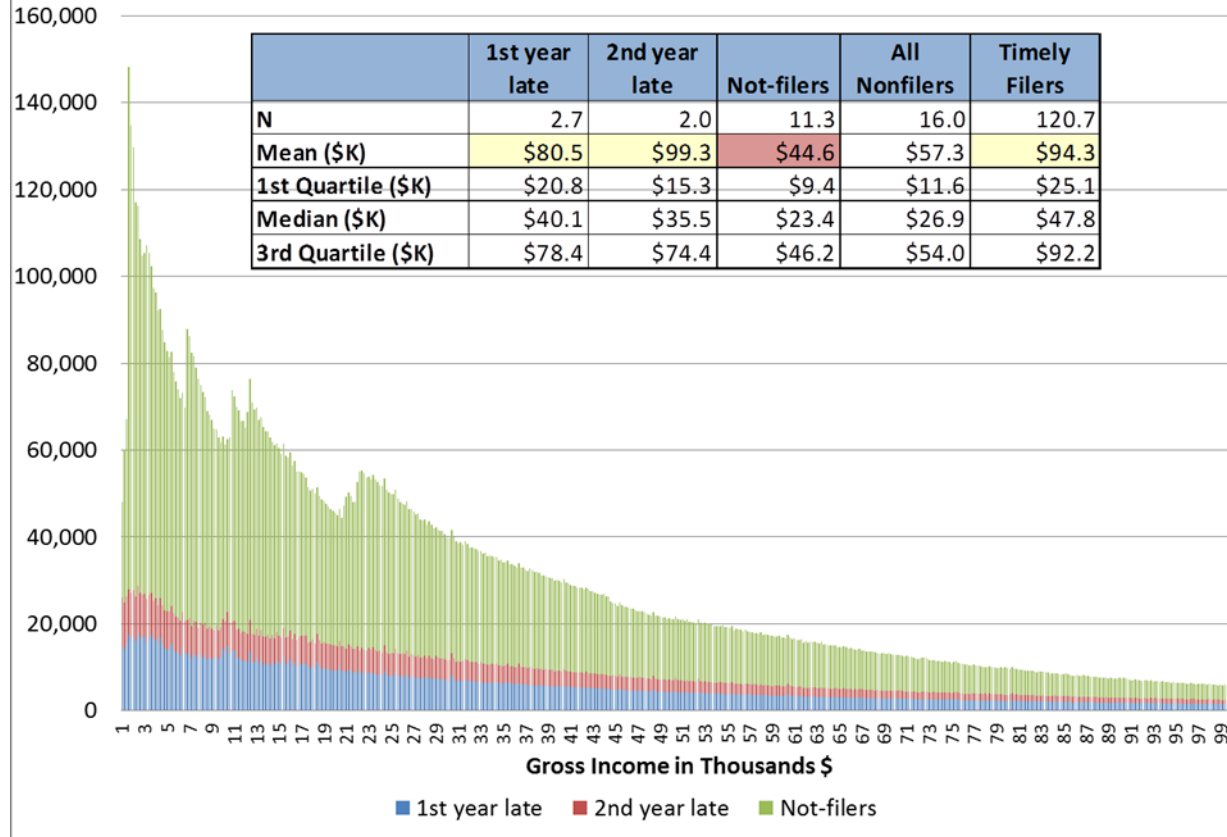
VFR by Primary Income Types (with SE Imputation), TY 2006 to TY 2014
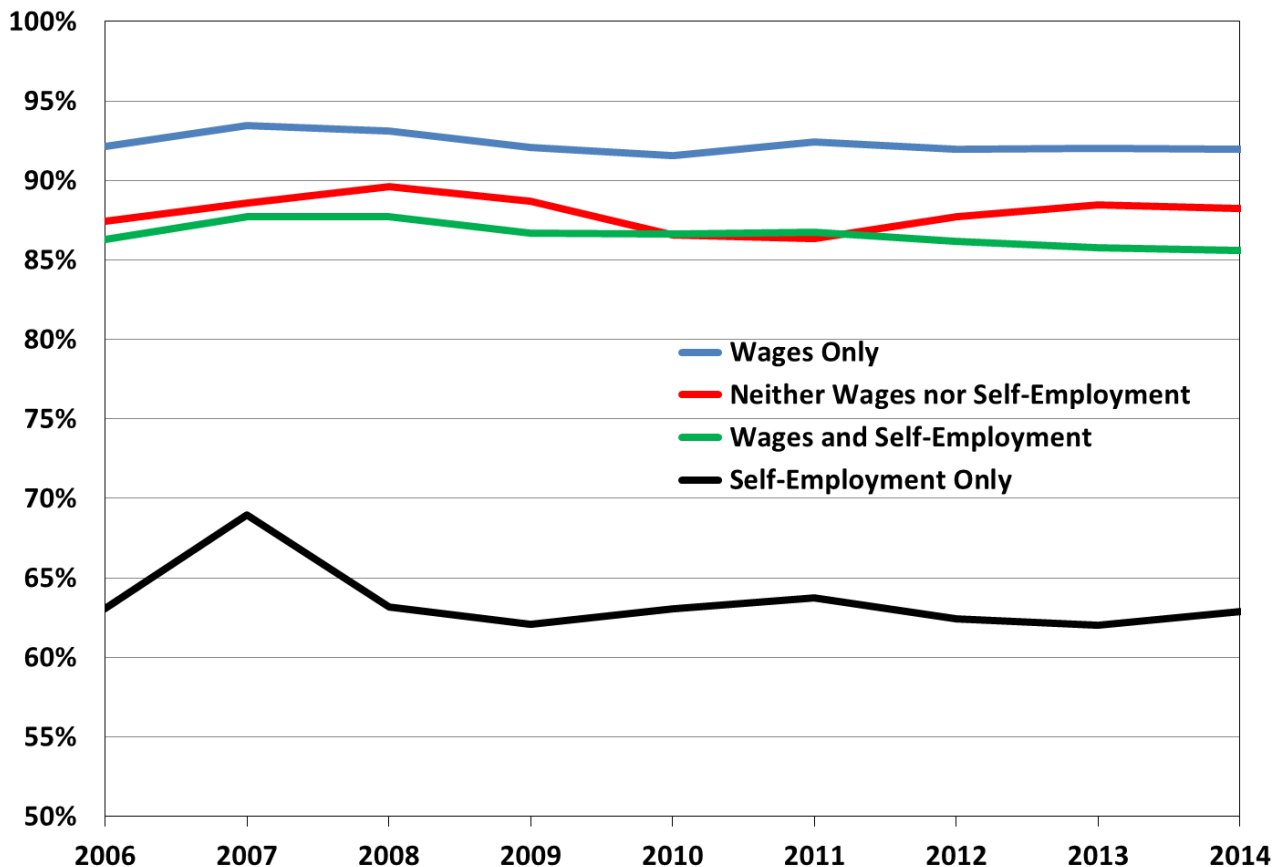
The VFR is much higher for those whose earned income is limited to wages and much lower for those with only SE income

Research, Applied Analytics, & Statistics

Distribution of Gross Income, Late and Not-Filers, TY 2014

	1st year late	2nd year late	Not-filers	All Nonfilers	Timely Filers
N	2.7	2.0	11.3	16.0	120.7
Mean ($K)	$80.5	$99.3	$44.6	$57.3	$94.3
1st Quartile ($K)	$20.8	$15.3	$9.4	$11.6	$25.1
Median ($K)	$40.1	$35.5	$23.4	$26.9	$47.8
3rd Quartile ($K)	$78.4	$74.4	$46.2	$54.0	$92.2

- **Distribution of gross income among nonfilers has a long tail**

- **Late filer and timely filer average gross income higher than for not-filers**

Future Work

- Impute corrected (single) filing status to some of those incorrectly claiming Head of Household status

- Improve imputation of tax units by drawing on information from prior year returns and SSA data

- Explore the use of expanded Census-IRS matched data to examine whether no-PIK records are disproportionately nonfilers.

- Explore use of IRS administrative data in multivariate analysis of drivers of nonfiling

Research, Applied Analytics, & Statistics