

# Does Encouraging Record Use for Financial Assets Improve Data Accuracy? *Evidence from Administrative Data*

Jonathan Eggleston and Lori Reeder

U.S. Census Bureau

March 20<sup>th</sup>, 2018

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau

# Audience Question #1

- Raise your hand if you have any of the following types of bank accounts:
  - Savings Account
  - Interest-Earning Checking Account
  - Money Market Account
  - Certificate Of Deposit (CDs)

# Audience Question #2

- Keep your hand raised if you know the answer to this question:
  - *How much interest income did you receive from your bank account between January 1st and the end of December 2017?*

# Assets & Measurement Error

- Problem: Many people are unaware of how much income they receive from their assets
- Problem could cause
  - Inaccurate data
  - Inaccurate measurement of income equality

# Assets & Measurement Error

- Potential solutions
  - Conduct interviews around tax season
  - **Encourage respondents to consult financial records**

# Record Use

- Benefits of encouraging record use unclear
  - Respondents who use records may already be aware of their finances
  - Experimental studies (Couper et al. 2013, Murphy et al. 2015): No significant effect of record use on measurement error

# Potential Limitations of Previous Experimental Studies

- Use indicators of measurement error, such as rounding, which may or may not be associated with actual measurement error
- Power Issues
  - Small sample size
  - Effect of experimental manipulation on record use was relatively small
  - Our estimates: Power around 10% in these experimental studies

# Our Study

- Compare survey to administrative data: More direct indicator of measurement error
- Novel estimation strategy: We use various measures of respondent motivation and precision to account for nonrandom selection
- Findings
  - Record use associated with a 21 to 43 percent decrease in measurement error
  - Record use increases interview length by 2.2%

# Data

- Administrative Data: IRS 1040 Tax Returns
- Survey Data: 2014 Panel of the Survey of Income and Program Participation (Wave 1)
- Variables on income from assets
  - Interest Income
  - Dividend Income
  - Gross Rental Income
- Sample Restrictions
  - Matched to IRS 1040
  - Positive (non-zero) asset income in both SIPP & IRS 1040

# Key Explanatory Variables

- Main variable: Indicator of consulting records for asset questions
  - 26% of respondents consulted records in 2014 SIPP (Wave 1)
  - Issue: respondents who consult records may be more engaged respondents who give accurate data, even if they didn't consult records

# Key Explanatory Variables

- Novel Method: Proxy variables for overall data accuracy
  - Respondent's average time per question
  - Respondent's item nonresponse rate for financial questions
  - Respondent's average amount of rounding for financial questions
- Key assumption: Proxy variables account for all confounding factors that are correlated with both
  - Measurement error in asset income
  - Record use

	Inverse Hyperbolic Sine (IHS) of Absolute Difference in Asset Income $\approx \log(\text{abs}(\text{SIPP}-\text{IRS}))$					
	Interest		Dividend		Rental	
Record Use	***-0.375 (0.049)	***-0.246 (0.053)	***-0.518 (0.103)	***-0.353 (0.105)	***-0.763 (0.262)	** -0.558 (0.274)
Time Per Question		-0.006 (0.005)		-0.006 (0.009)		-0.003 (0.016)
Item Nonresponse Rate		0.096 (0.115)		0.102 (0.199)		-0.050 (0.550)
Rounding Average		***.950 (0.138)		***1.072 (0.251)		***2.259 (0.809)
Proxy Variables	No	Yes	No	Yes	No	Yes
N. Obs. (Rounded)	7500	7500	2900	2900	1300	1300
Source: 2014 SIPP panel and 2013 IRS 1040 dataset						

	Inverse Hyperbolic Sine (IHS) of Absolute Difference in Asset Income					
	Interest		Dividend		Rental	
Record Use	***-0.375 (0.049)	***-0.246 (0.053)	***-0.518 (0.103)	***-0.353 (0.105)	***-0.763 (0.262)	** -0.558 (0.274)
Proxy Variables	No	Yes	No	Yes	No	Yes

Interest income:  
21.8% decrease in  
measurement error

Rental income:  
42.8% decrease in  
measurement error

Transformation:

$$\text{Percent decrease} = e^{(-0.246)} - 1 = 0.218$$

## Inverse Hyperbolic Sine (IHS) of Absolute Difference in Asset Income

Interest

Dividend

Rental

1% increase in average amount  
of rounding associated with a  
0.95% increase in  
measurement error.



Rounding Average

**\*\*\*.950**  
**(0.138)**

**\*\*\*1.072**  
**(0.251)**

**\*\*\*2.259**  
**(0.809)**

Proxy Variables

No

Yes

No

Yes

No

Yes

# Cost of Record Use

- Previous results: Potential benefit on data accuracy
- Cost: Record users may spend more time per question
  - Increase in respondent burden
  - Monetary cost for interviewer time
- Question: Does record use increase interview time, after controlling for other respondent characteristics?
  - Our contribution: First to control for confounding factors in a regression framework

# Regression for Seconds Spent Per Question for the Asset Income and Value Section

	Seconds Per Question
Record Use	<b>***3.499</b> (0.100)
Time Per Question in Other Sections (Seconds)	<b>***0.580</b> (0.014)
Interviewer's Time Per Question in Other Sections (Seconds)	<b>***0.310</b> (0.017)
Item Nonresponse Rate	<b>***-3.196</b> (0.190)
Rounding Average	<b>***-4.670</b> (0.195)
N. Observations (Rounded)	48,500
Source: 2014 SIPP panel	

# Regression for Seconds Spent Per Question for the Asset Income and Value Section

	Seconds Per Question
Record Use	***3.499 (0.100)



Record users spend an extra 3.5 seconds per asset question

# Regression for Seconds Spent Per Question for the Asset Income and Value Section

	Seconds Per Question
Record Use	***3.499 (0.10)



- Average number of asset value & income questions: 16.02
- Average length of personal interview: 41.65 minutes
  - Records users spend an extra 0.93 minutes in the asset section
  - Record use increases interview length by 2.2%

# Conclusion

- Our Study: First large scale comparison of survey to administrative data to determine whether record use is associated with improved data quality
- Record use associated with an 21 to 43 percent decrease in measurement error
- Record use associated with spending an extra 3.5 seconds on each asset question
  - Record use increases interview length by 2.2%

# Contact Information

## Jonathan Eggleston

Economist

Labor Force Statistics Branch

Social, Economic, and Housing  
Statistics Division

U.S. Census Bureau

Office: 301.763.2357

[jonathan.s.eggleston](mailto:jonathan.s.eggleston@census.gov)

[@census.gov](mailto:jonathan.s.eggleston@census.gov)

## Lori Reeder

Survey Statistician

Survey Improvement Research Branch

Social, Economic, and Housing  
Statistics Division

U.S. Census Bureau

Office: 301.763.9787

[lori.beth.reeder@census.gov](mailto:lori.beth.reeder@census.gov)

# Appendix

- Additional controls in regressions analyses
  - IRS asset income
  - Adjusted gross income
  - Indicators of imputed values
  - Percent of SIPP interest income which is imputed
  - Indicator of a proxy interview
  - Demographics variables such as age, race, and education
  - Interviewer's average time per question
  - Interview time of day and day of week
  - Number of assets owned
  - Employment income reported in SIPP

# Additional Regression Details

- The cross-sectional unit of analysis in this table is a tax unit (which is either a single person or a married couple)
- Sample consists of all individuals over age 15 and tax units with positive asset income in SIPP and IRS 1040
- In SIPP, IRS data is not used to impute asset data

# Time Means

Section of Personal Interview	Record Use	
	Yes	No
Asset Value and Income	19.5 (0.14)	12.8 (0.04)
All Other Sections	14.7 (0.07)	12.6 (0.04)

# Matching Algorithm

