Research, Analysis & Statistics

STATISTICS OF INCOME

Welcome
Statistics of Income
2015 Consultants Panel Meeting

Barry Johnson
Director Statistics of Income Division

Thanks to the Committee on National Statistics for hosting

Special thanks to Dr. Connie Citro, Director of the Committee on National Statistics, National Academy of Science

A few logistics

- Facilities
- Lunch
- Travel Vouchers (please see Wanda Robinson)
- Thanks for help organizing meeting:
  - SOI Staff: Brian Balkovic, Paul Bastuscheck, Melissa Belvedere, Rose Defalco, Martha Harris, Wayne Kei, Melissa Ludlum, Clay Moulton, David Paris, Wanda Robinson, Georgette Walsh
  - NSF Staff: Eileen LeFurgy, Jesse Willis
- If you need anything during the day, please see Brian Balkovic, Rose Defalco, or Wanda Robinson

SOI Panel Members

Michael Allen

Rosanne Altshuler

Jenny Bourne

Peter Brady

Len Burman

Martin David

Daniel Feenberg

John Graham

Julia Lane

Jim Nunns

George Plesko

Fritz Scheuren

Lin Smith

Robert Strauss

Michael Udell

Patricia Whitridge (guest panelist)

Updates

RAS

- Rosemary Marcuss retired, Alain Dubois is now acting Director
- Research, Analysis and Statistics/Office of Compliance Analytics redesign

SOI

New or improved products detailed on preread

- Resources
- Communications strategy
  - Web redesign efforts
  - Infographics
  - Updated *SOI Bulletin* articles
- .Net conversion
- 5-Year Business Plan
  - Interviews with SOI staff and internal stakeholders
  - Interviews with external stakeholders
- Acting Director July 20-October 2

2015 Consultants Panel Agenda

What's New at SOI?		
Welcome	Barry Johnson	9:00 am
Joint Statistical Research Program Update	Mike Weber	
A New Approach to Producing Migration Data	Kevin Pierce	
Discussion	Panel	
Break		**9:55 am**
Thinking Big About SOI Data		
Big SOI	Fritz Scheuren	10:05 am
Partnership Data Research	Danny Yagan, Eric Zwick, Owen Zidar	
What is SOI's value added	Jim Nunns	
AAPOR Big Data Report	Julia Lane	
Canada Revenue Agency's Business Intelligence Strategy and Agency Data Program	Patricia Whitridge	
LEI update and session discussion Discussion	Arthur Kennickell Larry May, Ralph Rector. Panel	
Lunch		**12:05 pm**

2015 Consultants Panel Agenda (continued)

Are Piketty and Zucman Getting it Right? Evaluating Distributional Statistics Based on Aggregate Data		
More Than They Realize: The Income of the Wealthy and the Piketty Thesis	Jenny Bourne	1:20 pm
Measuring Income at the Top	John Sabelhaus	
Mortality Differentials - How Much Longevity Can Money Really Buy?	Brian Raub	
Discussant	Len Burman	
Discussion	Panel	
A Productive Partnership, Joint Work with Stanford	David Grusky	2:30 pm
Discussion	Panel	
An Overview of the SOI Consultants Panel	George Plesko	3:10 pm
Discussion	Panel	
Adjorn		

2014 SOI JOINT STATISTICAL RESEARCH PROGRAM

2014 Joint Statistical Research Program

- 87 Proposals were submitted (43 from grad students)
- 13 focused on compliance
- 13 focused on corporate tax issues
- 57 focused on individual tax issue
- 10 focused on a combination of the two
- Over 20 proposals required direct matching to outside data sets that raised legal and privacy concerns
- Several papers required data that are not available
- At least 12 did not credibly connect research to tax administration

2014 Joint Statistical Research Program

- 12 proposals were approved
  - 7 Tax Policy focused proposals (to be conducted by SOI)
  - 5 Compliance focused proposals (to be conducted by the Research Division)

Approved Projects (SOI)

Expanding SOI Data Products on Flow-Through Entities

Joseph Rosenberg - Tax Policy Center

James Nunns - Tax Policy Center

Tax incentives and changes in labor and capital income inequality

Wojciech Kopczuk – Columbia University

Approved Projects (SOI)

Effect of Estate Tax on Wealth Accumulation, Labor Supply, and Cross-State Migration

Jon Bakija - Williams College

Income Risk in the United States, and the Effectiveness of Insuring through Labor Supply, College Education, Assortative Marriage, and Federal Taxes and Transfers

James J. Heckman – University of Chicago

Magne Mogstad, - University of Chicago

Bradley Setzler - University of Chicago

Approved Projects (SOI)

A Protocol for Classifying the Taxpayer's Occupation

David B. Grusky - Stanford University

Michael Hout - New York University

David Johnson - Bureau of Economic Analysis

Michelle Jackson - Stanford University

Jonathan Fisher - Stanford University

Pablo Mitnik - Stanford University

Approved Projects (SOI)

The Impact of Income Volatility on Measured Cross-Section Income Inequality

Jeffrey P. Thompson - Federal Reserve Board

John Sabelhaus - Federal Reserve Board

Distribution of Tax Expenditures from a Permanent Income Perspective

Katharine Abraham - University of Maryland

Approved Projects (Research)

Estimating the Causal Effect of Third-Party Reporting on Small-business Tax Compliance

James Alm – Tulane University

Bibek Adhikari – Tulane University

Nonprofit Taxable Activities - How and which nonprofit organizations use taxable revenues to supplement other revenue streams within their operations

Steven Balsam – Temple University

Eric Harris – Rutgers University

Approved Projects (Research)

(Non)disclosure of subsidiary locations and corporate tax behavior

Scott Dyreng - Duke University

Jeff Hoopes - Ohio State

Jaron Wilde - University of Iowa

The Effect and Effectiveness of Tax Auditors

Joel Slemrod - University of Michigan

Ugo Troiano - University of Michigan

Shlomo Yitzhaki - Hebrew University of Jerusalem

Approved Projects (Research)

The Impact of the Offshore Voluntary Disclosure

Joel Slemrod - University of Michigan

Jeffrey Hoopes - Ohio State University

Daniel Reck - University of Michigan

SOI MIGRATION DATA: A NEW APPROACH

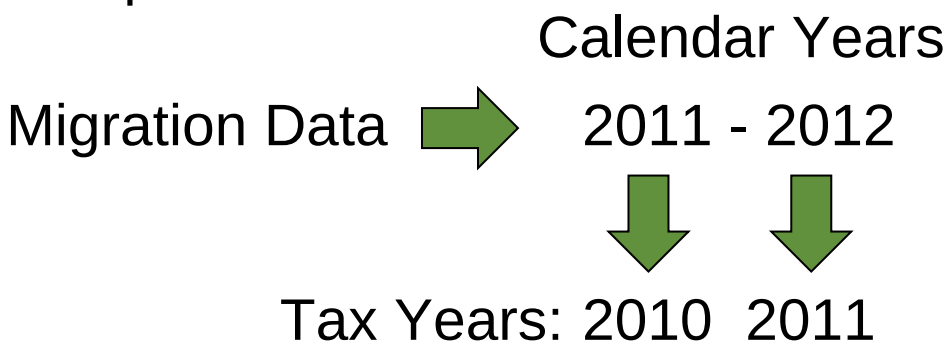
What are the Migration Data?

Migration data show the movement of individuals, via the address listed on Form 1040, over a two-year period

What are the Migration Data?

Migration data show the movement of individuals, via the address listed on Form 1040, over a two-year period

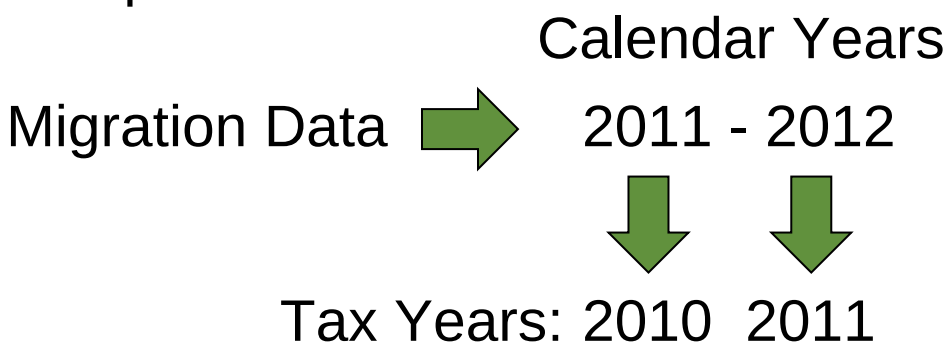
As an example…



What are the Migration Data?

Migration data show the movement of individuals, via the address listed on Form 1040, over a two-year period

As an example…



The data are available at the State or county level:

a. Number of inflows – residents moving in
b. Number of outflows – residents moving out

What are the Migration Data?

State level

(1) State-to-State Inflow

(2) State-to-State Outflow

County level

(1) County-to-County Inflow

(2) County-to-County Outflow

What are the Migration Data?

State level

(1) State-to-State Inflow

(2) State-to-State Outflow

County level

(1) County-to-County Inflow

(2) County-to-County Outflow

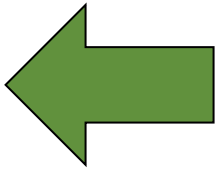
Origin from Alabama (State Code)	Destination into			Number of returns	Number of exemptions	Adjusted gross income (AGI)
	State Code	State	State Name	(1)	(2)	(3)
01	96	AL	AL Total Migration US and Foreign	51,971	107,304	2,109,108
01	97	AL	AL Total Migration US	50,940	105,006	2,059,642
01	98	AL	AL Total Migration Foreign	1,031	2,298	49,465
01	01	AL	AL Non-migrants	1,584,665	3,603,439	87,222,478

Previous Migration Data and New Migration Data

Previous

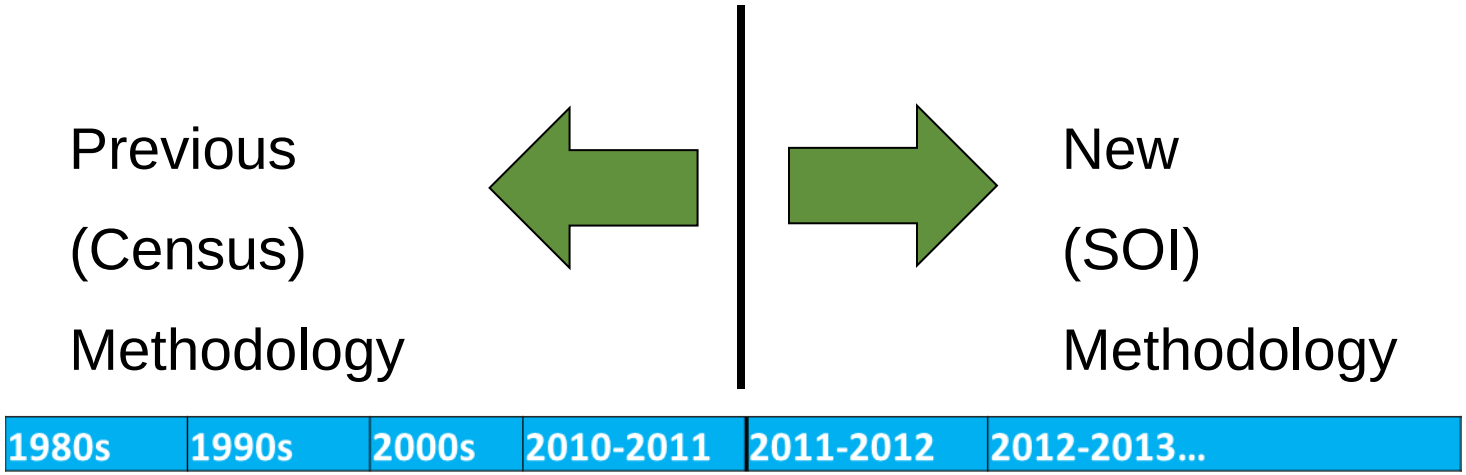
(Census)

Methodology

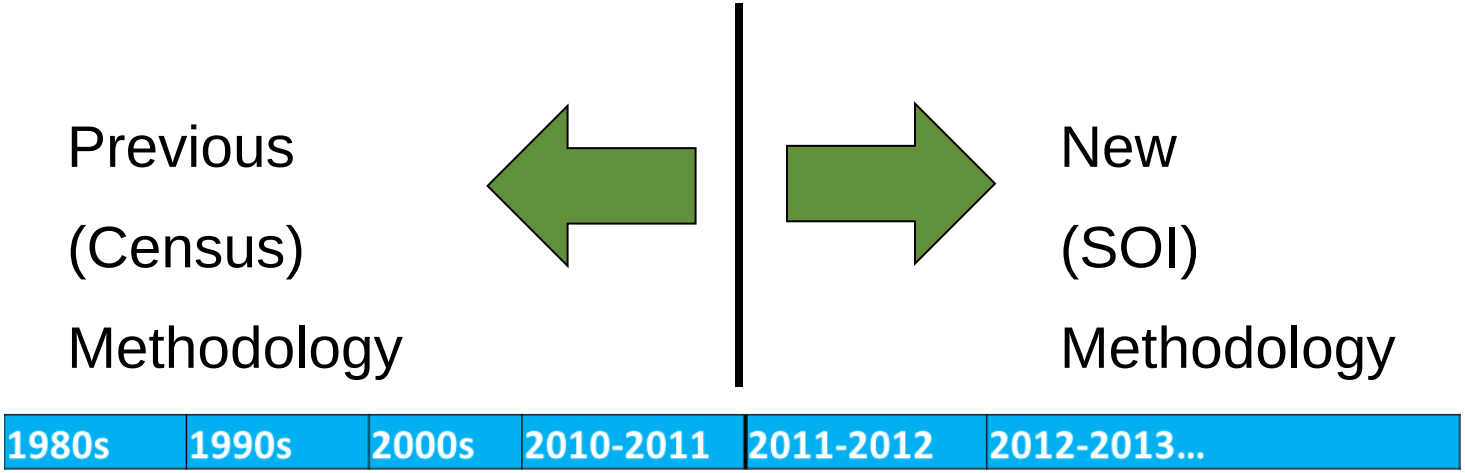


1980s	1990s	2000s	2010-2011	2011-2012	2012-2013…

Previous Migration Data and New Migration Data

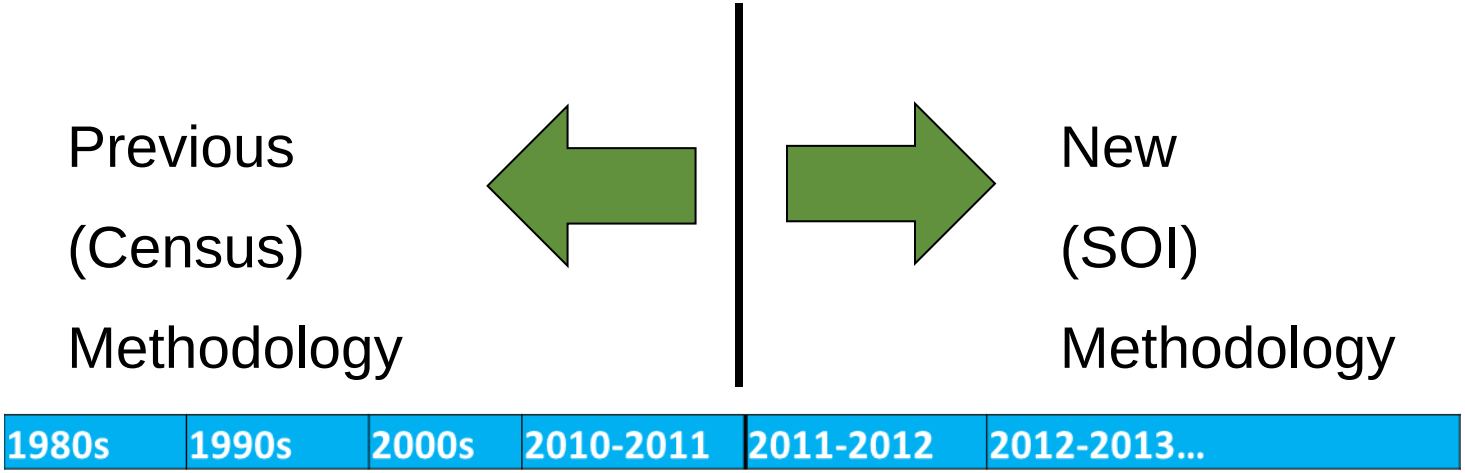


Previous Migration Data and New Migration Data



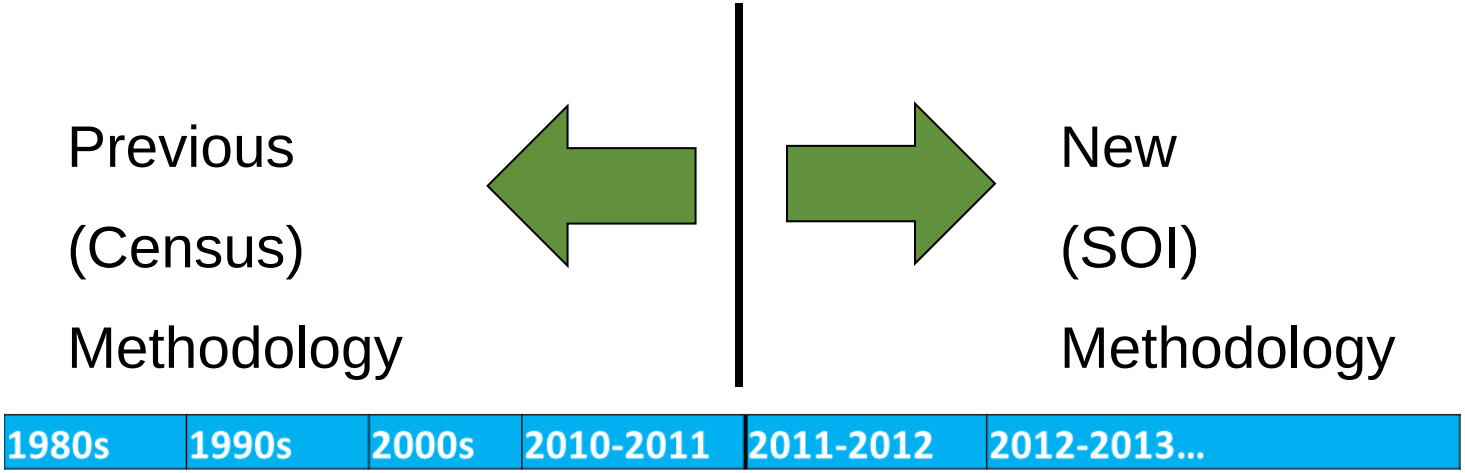
- Based on partial-year data

Previous Migration Data and New Migration Data



- Based on partial-year data
- Based on full-year data

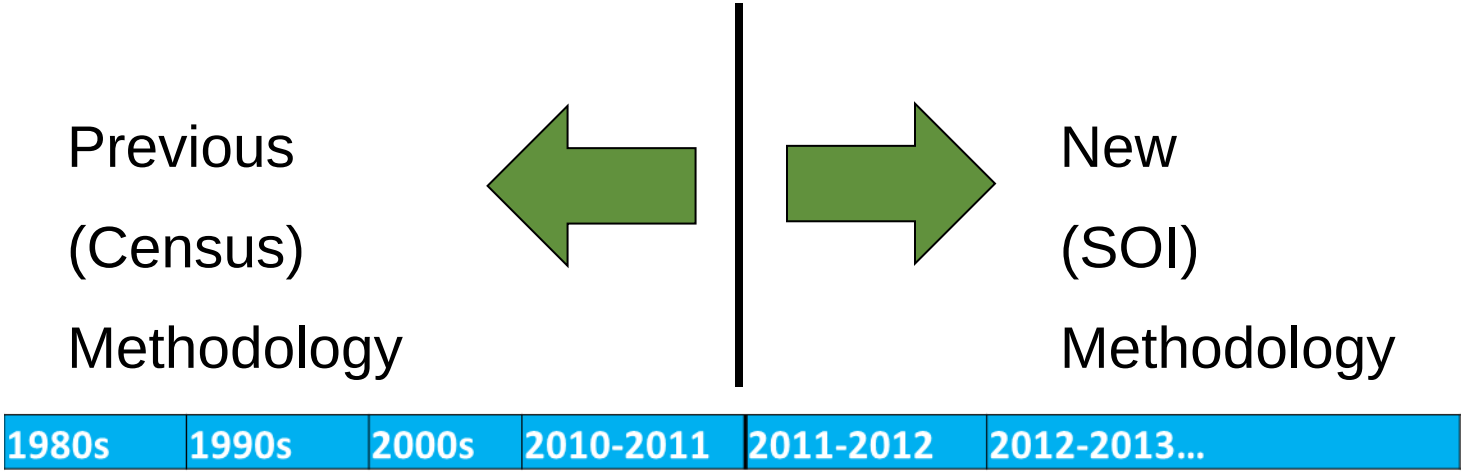
Previous Migration Data and New Migration Data



- Based on partial-year data
- Matched on primary TIN

- Based on full-year data

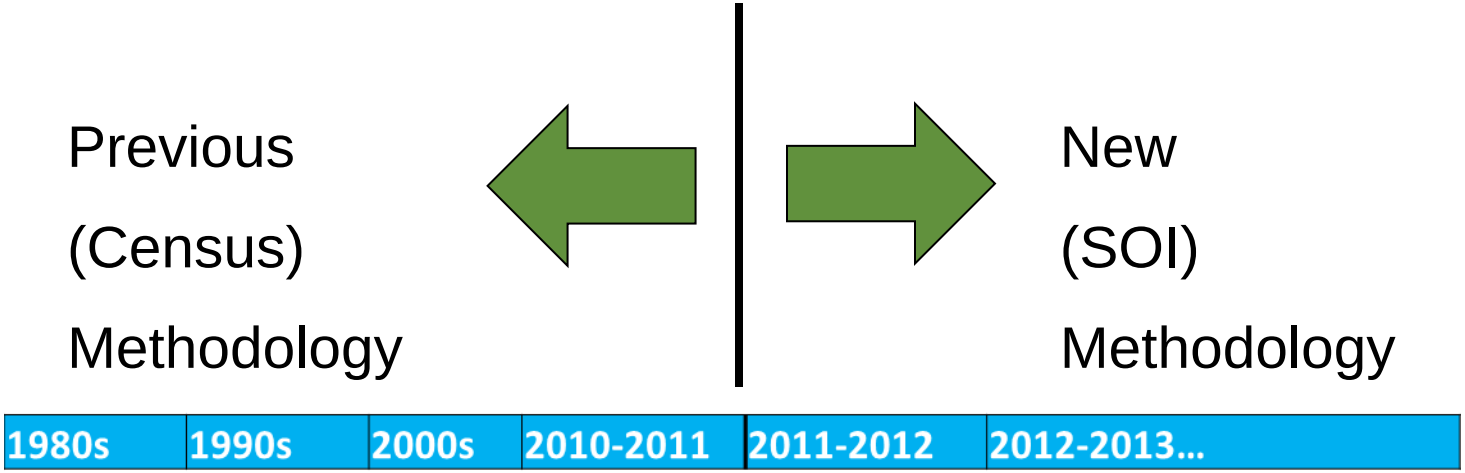
Previous Migration Data and New Migration Data



- Based on partial-year data
- Matched on primary TIN

- Based on full-year data
- Matched on primary, secondary, and dependent filer TINs

Previous Migration Data and New Migration Data

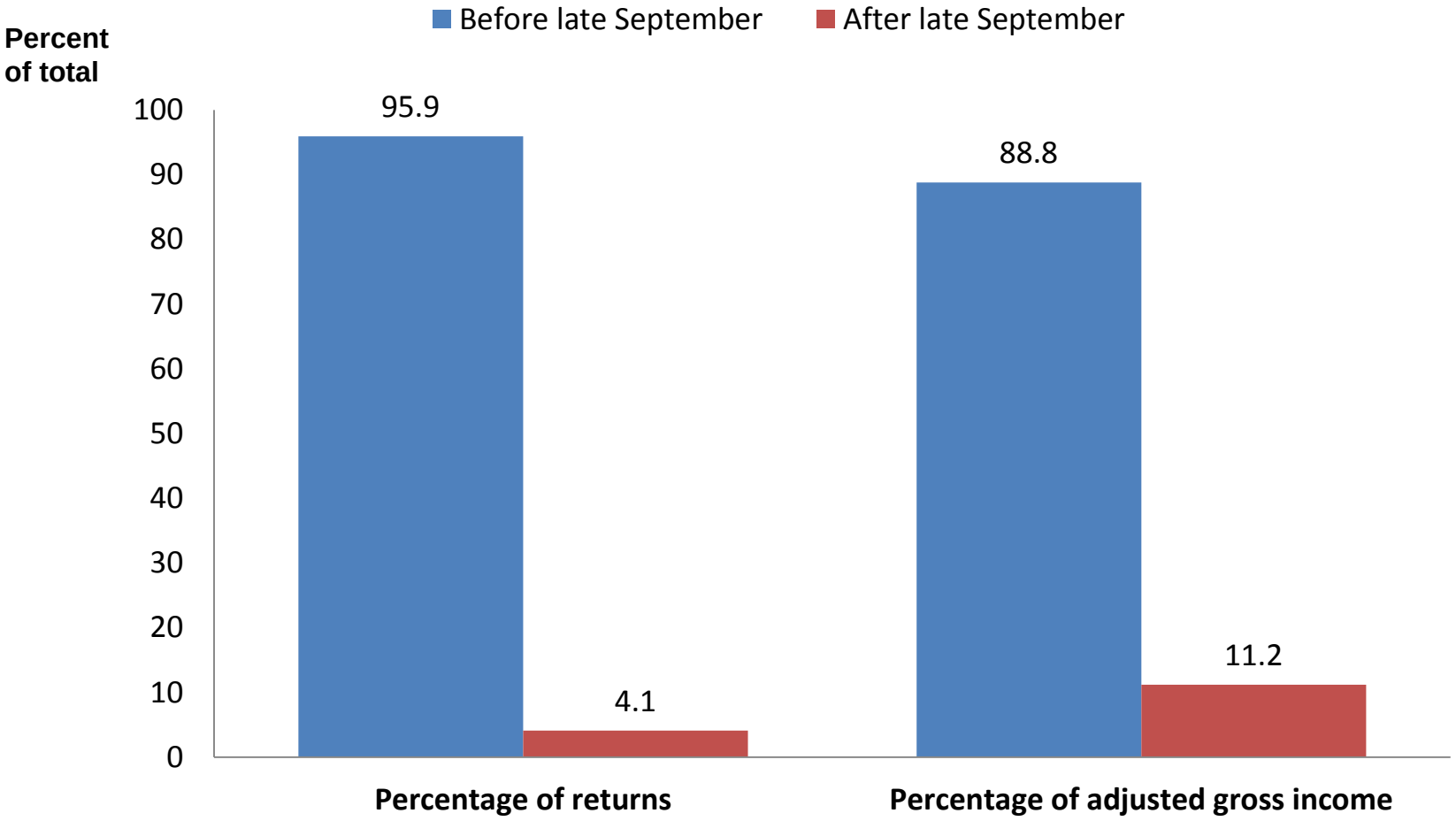


- Based on partial-year data
- Matched on primary TIN

- Based on full-year data
- Matched on primary, secondary, and dependent filer TINs
- Inclusion of summary flows by AGI and age of the primary taxpayer

FROM PARTIAL-YEAR DATA TO FULL-YEAR DATA

Percentage of Returns and AGI received Before and After late September, Calendar Year 2012



EXPANDED YEAR-TO-YEAR RETURN MATCHING

Expanded Matching between filers

YEAR 1	YEAR 2	Percent of the total matched returns
1. Primary filer	→Primary filer	

Expanded Matching between filers

YEAR 1		YEAR 2	Percent of the total matched returns
1. Primary filer	→	Primary filer	
2. Primary filer	→	Secondary filer	
3. Secondary filer	→	Primary filer	
4. Secondary filer	→	Secondary filer	
5. Dependent filers	→	Primary filer	
6. Dependent filers	→	Secondary filer	

Expanded Matching between filers

YEAR 1		YEAR 2	Percent of the total matched returns
1. Primary filer	→	Primary filer	94.6%
2. Primary filer	→	Secondary filer	
3. Secondary filer	→	Primary filer	
4. Secondary filer	→	Secondary filer	
5. Dependent filers	→	Primary filer	
6. Dependent filers	→	Secondary filer	

Expanded Matching between filers

YEAR 1		YEAR 2	Percent of the total matched returns
1. Primary filer	→	Primary filer	94.6%
2. Primary filer	→	Secondary filer	0.8%
3. Secondary filer	→	Primary filer	1.7%
4. Secondary filer	→	Secondary filer	less than 0.1%
5. Dependent filers	→	Primary filer	2.8%
6. Dependent filers	→	Secondary filer	less than 0.1%

Expanded Matching between filers

YEAR 1	YEAR 2	Percent of the total matched returns
1. Primary filer	→ Primary filer	94.6%
2. Primary filer	→ Secondary filer	0.8%
3. Secondary filer	→ Primary filer	1.7%
4. Secondary filer	→ Secondary filer	less than 0.1%
5. Dependent filers	→ Primary filer	2.8%
6. Dependent filers	→ Secondary filer	less than 0.1%

THE GROSS MIGRATION FILE

Gross Migration File

Migration Flows ⇨	Total Matched Returns (1)	Non-migrant Returns (2)	Outflow Returns (3)	Inflow Returns (4)	Same State Returns (5)

Gross Migration File

Migration Flows ⇨

Total Matched Returns (1)	Non-migrant Returns (2)	Outflow Returns (3)	Inflow Returns (4)	Same State Returns (5)

⇩

Age Categories

Under 26 (1)	26 under 35 (2)	35 under 45 (3)	45 under 55 (4)	55 under 65 (5)	65 and over (6)

Gross Migration File

Migration Flows

Total Matched Returns (1)	Non-migrant Returns (2)	Outflow Returns (3)	Inflow Returns (4)	Same State Returns (5)

Age Categories

Under 26 (1)	26 under 35 (2)	35 under 45 (3)	45 under 55 (4)	55 under 65 (5)	65 and over (6)

State

Alabama
Alaska
Arizona
Arkansas
California
….
Wyoming

Gross Migration File

Migration Flows

Total Matched Returns (1)	Non-migrant Returns (2)	Outflow Returns (3)	Inflow Returns (4)	Same State Returns (5)

Age Categories

Under 26 (1)	26 under 35 (2)	35 under 45 (3)	45 under 55 (4)	55 under 65 (5)	65 and over (6)

State

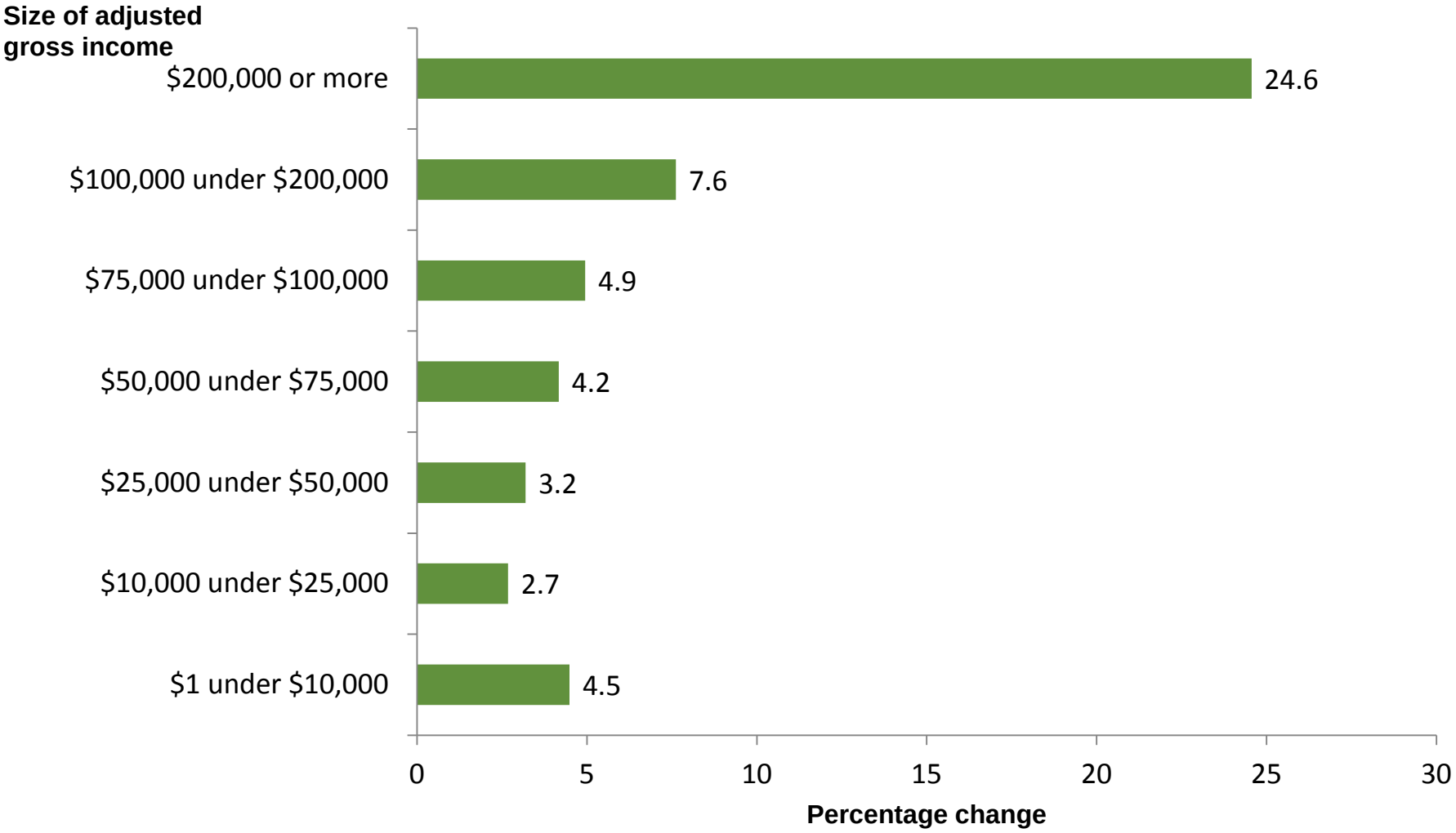
Alabama
Alaska
Arizona
Arkansas
California
….
Wyoming

AGI Categories

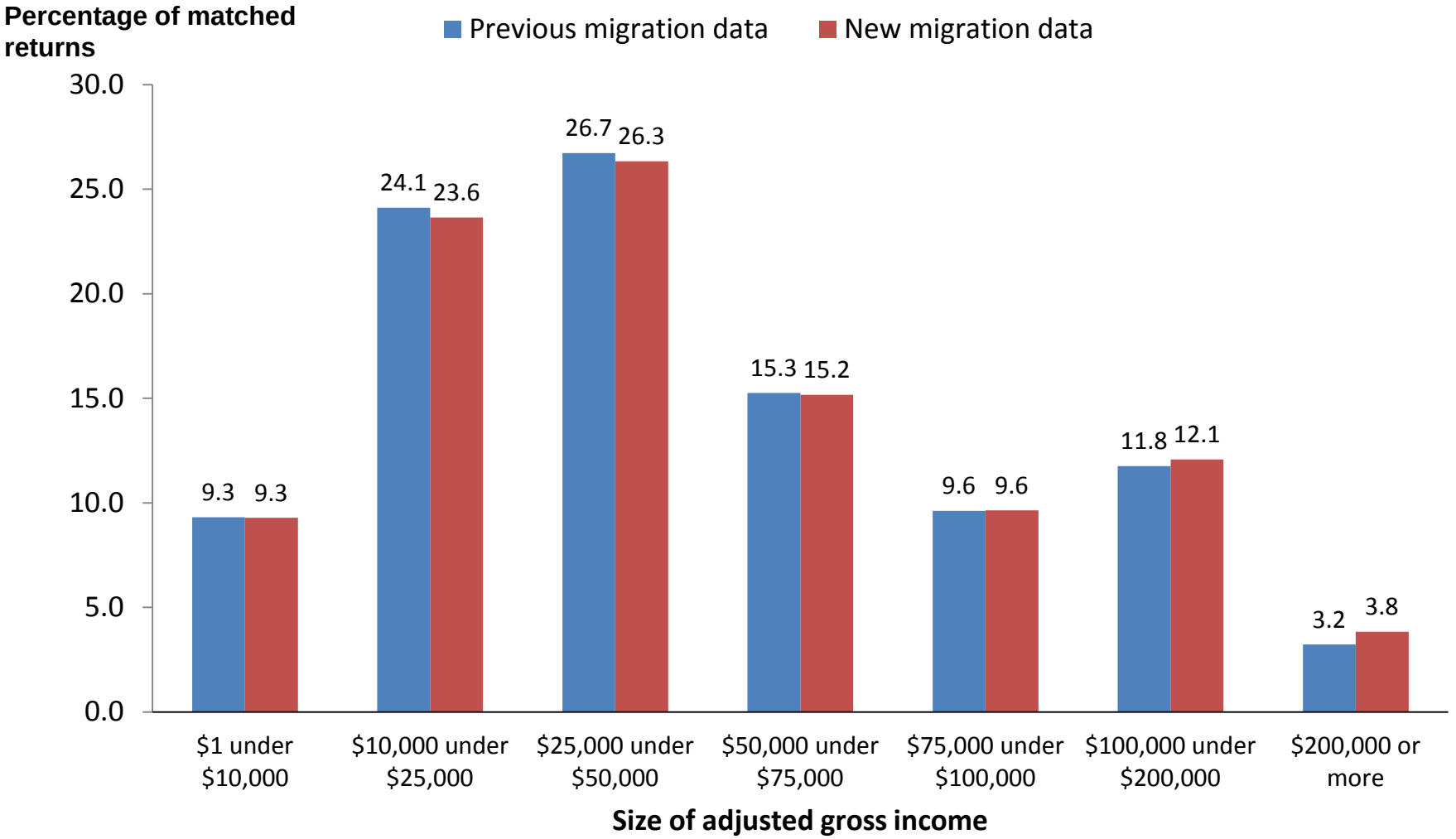
$1 under $10,000
$10,000 under $25,000
$25,000 under $50,000
$50,000 under $75,000
$75,000 under $100,000
$100,000 under $200,000
$200,000 or more

COMPARING OLD VS. NEW MIGRATION DATA

Percentage Change in Number of Returns for SOI Migration Data, by AGI, Calendar Years 2011-2012



Distribution of Matched Returns, by AGI, Calendar Years 2011-2012



Net-Migration Rate

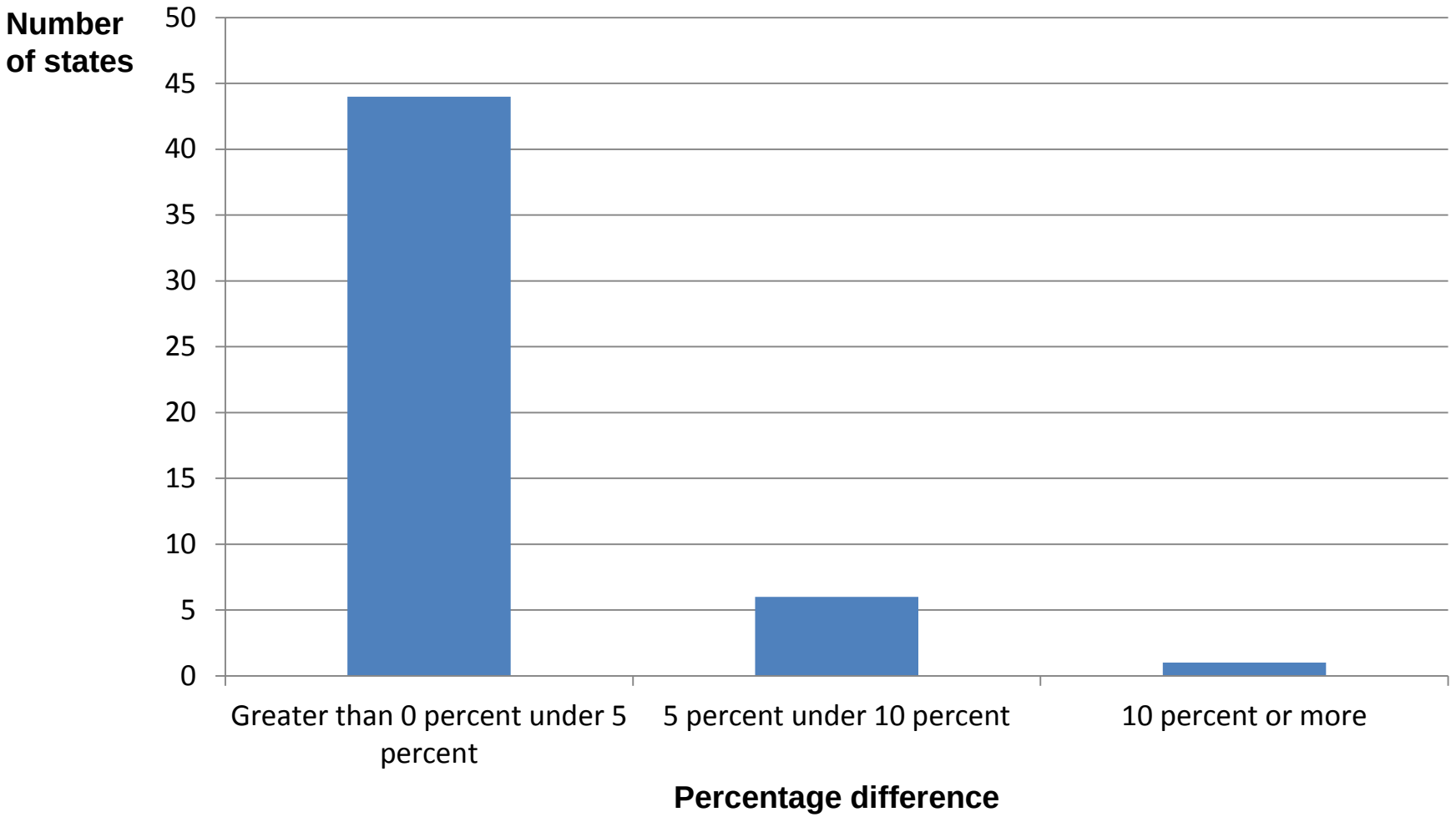
$$\text{Net-Migration Rate} = \frac{\text{In-Migrants} - \text{Out-Migrants}}{(\text{Non-migrants} + \text{Out-Migrants})}$$

Net-Migration Rate

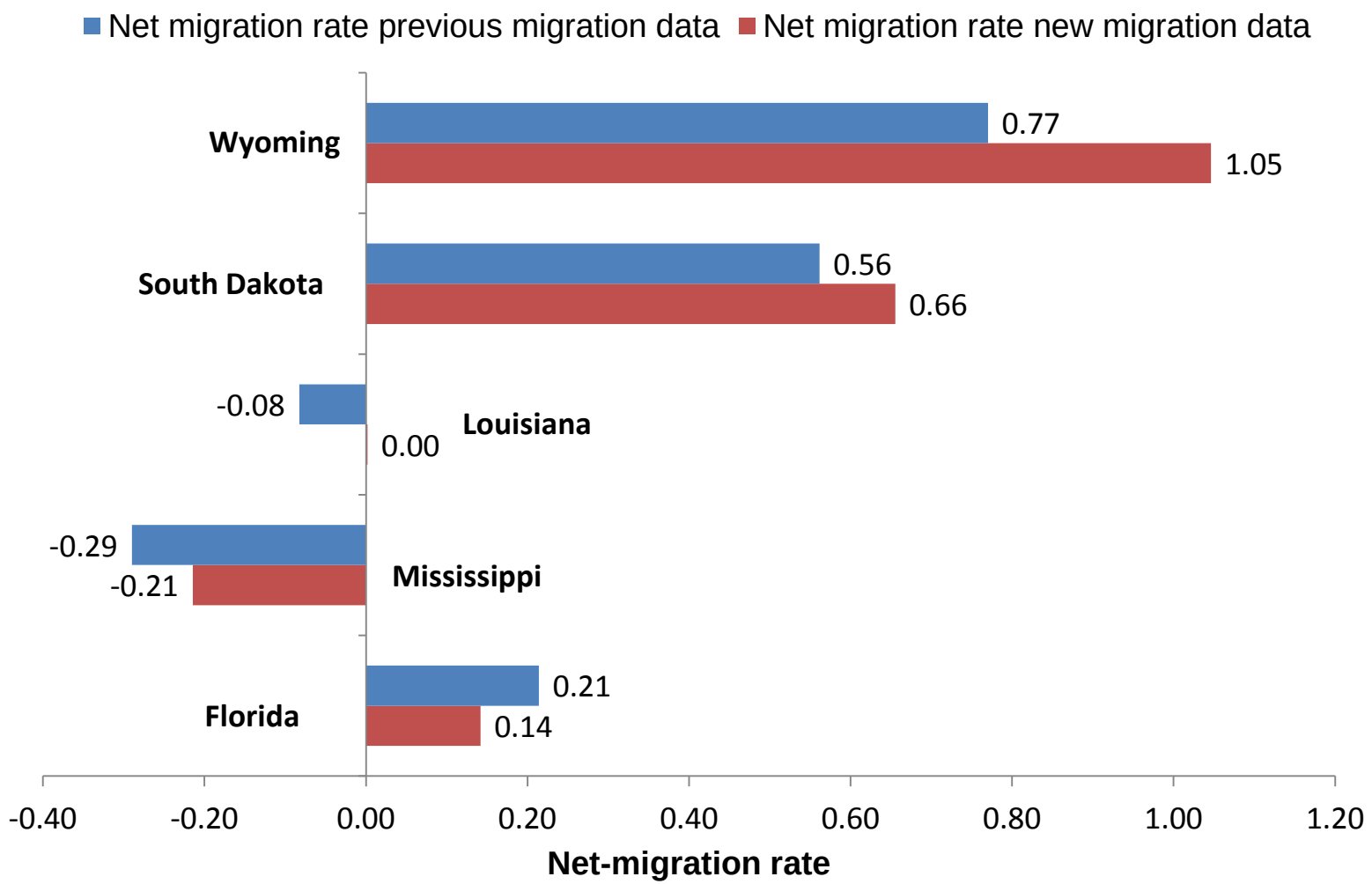
$$\text{Net-Migration Rate} = \frac{\text{In-Migrants} - \text{Out-Migrants}}{(\text{Non-migrants} + \text{Out-Migrants})}$$

- Computed for all 50 States, plus DC
- Previous and new migration data

Number of States, by Percentage Difference of the Net-Migration Rates of Old and New Data



Top 5 States with Largest Net-Migration Rate Differential, Calendar Years 2011-2012



Panel Discussion

What's New At SOI?

Discussion Question

- Are we doing enough to ease the transition to the new migration data series, or is there something else that would be useful?

Break:
10 minutes

Next:
Thinking Big About SOI Data

THINKING BIG ABOUT SOI DATA

BIG SOI

Fritz Scheuren
NORC at the University of Chicago

SOI-O-O-O BIG!!!

- Time to transform SOI again
- Into the new Big Data World
- Now visible to all!!!
- Title taken from a 1924 Edna Ferber Novel
- Of course, the “I” was put in
- To give it an SOI connection!!!

Big Data as Challenge?

- SOI must change drastically to survive
- To doubt that is to be asleep!!!
- But the change is one SOI has been getting ready for
- All SOI has to do is want to change more/bigger

Big Data as Opportunity?

- Opportunity or Adventure?
- Planning is the Difference?
- Commit NOW then do!
- One small step at a time!
- Pilot and then pilot again?
- But keep moving, fast!!!
- The field is full of other players
- Keep a Wise head and young heart!

The five Eras of SOI so far?

- The Pre SOI Age at IRS
- The Pre Electronic SOI Age
- Golden Age of SOI Statistics
- Golden Age of SOI “Data”
- Big SOI Metadata Age is now
- Two Analytic Program Examples

The Pre SOI Age

- SOI was part of a General Reform of Government
- Greater transparency was part of that movement
- Statistical Summaries of the tax system were needed

The Pre Electronic SOI Age

- SOI was the only major source of Tax Return Information until the 1950/60s
- SOI Statistics for individuals and corporations were annual then
- But corporations were not even sampled until the 1950s

Golden Age of SOI Statistics

- At the beginning SOI customers only could use and, hence, only wanted SOI Statistics
- Of course, more timely statistics and always in more detail
- Special SOI studies were an answer for a while

But Change was Occurring

- The 1960s were changing America, mostly for the better IRS/OTA too
- The IRS Master File was improving and SOI's clients were upgrading their computer systems, like now, sadly faster that SOI
- This led to a demand for SOI Data over SOI Statistics/Analytics

Golden Age of SOI “Data”

- SOI expands its program reach beyond IRS enumerative samples
- Linkages were strengthened with SSA and FRB/Census survey data
- But still very difficult and expensive and ad hoc

Why Not More?

- As successful as SOI was it did not adapt/adopt fast enough to the growing data dense world
- Is SOI committed to really major changes now. It must be?
- This is the time to break free of the parts of the past where SOI was kept back and “Run to Glory”

What is Big Data?

- “Big Data” buzz has everyone looking at new options, SOI too
- But what is “Big Data” really?
- No fixed definition yet
- I’ll use Rob Kitchin’s definition
- And examples from June Journal

Defining SOI in Big Data Era

- Big Data is Big N and Big P?
- Big N (number of units) is way more than typically was used just a few years ago
- Making P (number of items per unit) large is still hard to come by and a place for record linkage

SOI Going BIG Little by Little

- Barriers abound bureaucratic, statistical, and financial
- But SOI has already started and needs to expand relentlessly
- SOI can walk the way ahead!!!
- An Inter/intra agency respect-respect approach is needed

Elements on SOI “Going Big”

- Building SOI Data Relationship across Agencies
- Handling Confidentiality Issues which grow with linkage
- Attempting to Measure increases in Linkage uncertainty

Confidentiality in Linkage Issues

- Very Hard at every stage
- Merging, Matching, Sharing Results
- Several small-scale precedents
- No routine practice exists
- BLS/Census Synthetic Data Project mentioned last year offers hope!

Measuring Linkage Uncertainty

- Again, very hard at every stage
- Linkage costs unaffordable, even undoable without errors
- Trading Sampling Error for Big Data linkage errors?
- May be a good trade but it depends? Sometimes not!

Try anyway! Experience will tell?

- The sub-systems to be “bigged” were not designed to be linked
- These are statistical problems and can be satisficed/if not solved!
- Beyond our pay grade are the turf problems but once people work together more, who knows?

Two Partial Big Data Examples

- Individual and organizational examples chosen -- one each
- The “Shock of Recognition” – We have been “bigging” for a while
- Still more than just scaling up?
- Two “bigging” to Start SOI with!

Individual's Little "Bigging"

- Medicare Public Use(PUF) case
- SCF/CPS expanding the larger sample over years/record types
- Expanding the Main SOI public sample as in Medicare Example
- Synthetic Expansion of SCF/CPS data too larger SOI data sets

Partnership's Little "Bigging"

- Partnering with OTA to further the BLS linkages with the unemployment Tax System
- Already broadly covered by same IR Code as SOI
- Census/BLS Synthetic precedent exists and could fit nicely

When to Select What Sample?

- SOI Partnership Program goals:
- Editing/interpreting dollar totals?
- Studying business demography?
- Partnership income recipients?
- Flexible cross-section and longitudinal tools, new and old!!!
- Before/after linkages up and down?

Bless You's and Thanks to All

- First to Barry who has allowed me to speak
- Then to the Panel that has stood by SOI all these years
- Then to my sisters and brothers at SOI my IRS family

Joint Business Tax Data Project: Lessons from Estimating Average Partnership Tax Rates

Danny Yagan, UC Berkeley and NBER
Owen Zidar, Chicago Booth and NBER
Eric Zwick, Chicago Booth and NBER

June 2015

General need to link firms to ultimate owners

- OTA need: Estimate tax rate and revenue consequences of tax reform
- Growing reality: Business taxation occurs at owner level, after flowing through ownership chains
- Pilot: Link partnerships to ultimate owners to estimate current average tax rate on partnership income

- This presentation:

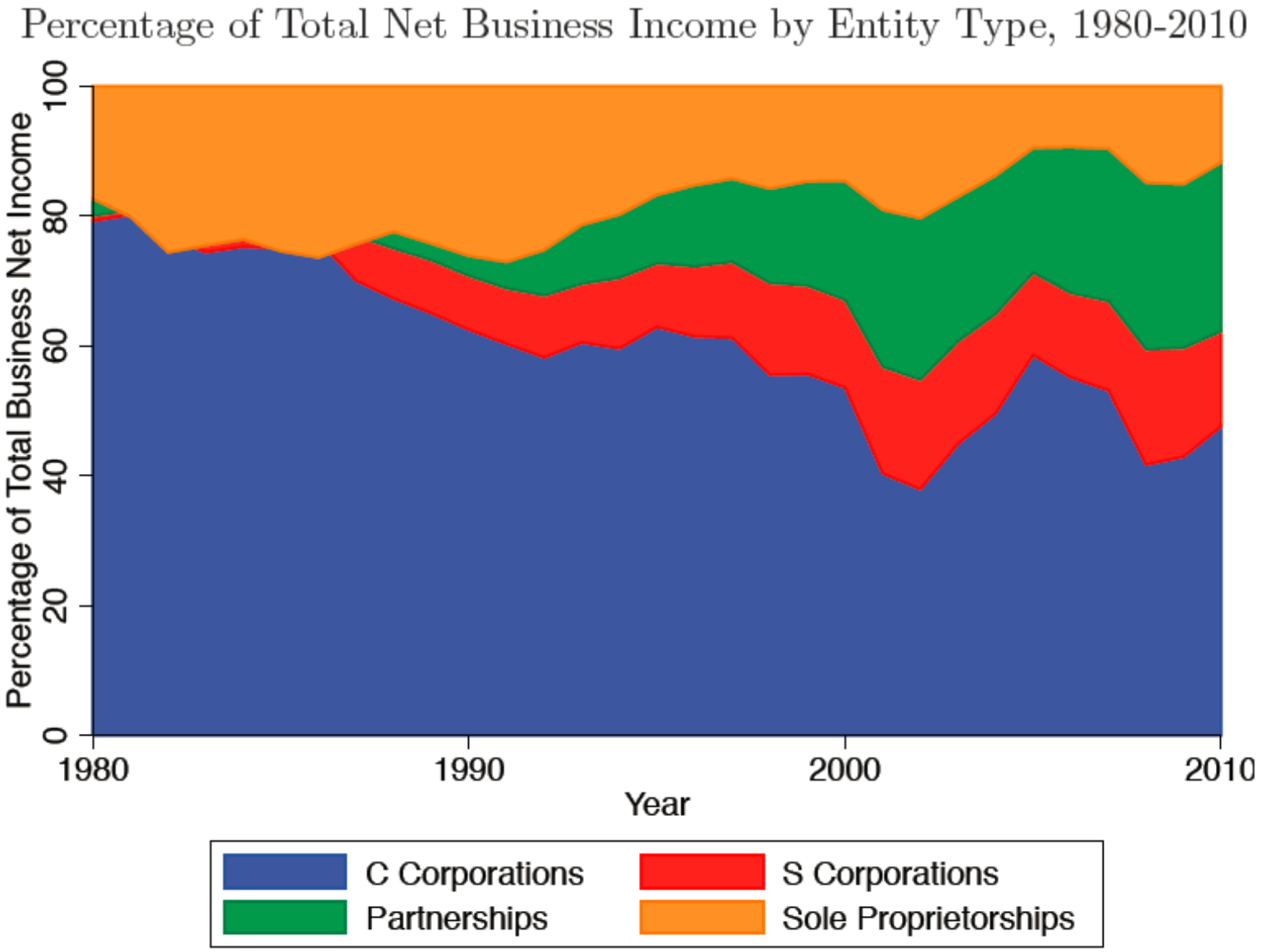
  1. Our partnership project and findings
  2. Two actionable suggestions for SOI work going forward

Partnerships in the United States: Who Owns Them and How Much Tax They Pay

Michael Cooper, U.S. Treasury Department
John McClelland, U.S. Treasury Department
James Pearce, U.S. Treasury Department
Richard Prisinzano, U.S. Treasury Department
Joseph Sullivan, U.S. Treasury Department
Danny Yagan, UC Berkeley and NBER
Owen Zidar, Chicago Booth and NBER
Eric Zwick, Chicago Booth and NBER

May 2015

Motivation: What does this transformation mean?



Source: DeBacker-Prisinzano (2015), Petska-Parisi-Luttrell-Davitian-Scoffie (2005)

This presentation

- Business activity lies increasingly outside of the C-corporate sector
- Tax reform requires facts on tax rates for each sector

- We provide:

  1. Linked partnership-partner data
  2. Facts on the distribution of business income
  3. Underlying tax rates for the partnership sector

Part 1: Partnerships file a business inc. tax return...

Form **1065**
Department of the Treasury
Internal Revenue Service

U.S. Return of Partnership Income

For calendar year 2011, or tax year beginning _____, 2011, ending _____, 20_____.
▶ See separate instructions.

OMB No. 1545-0099
2011

A Principal business activity	**Print or type.**	Name of partnership	**D** Employer identification number
B Principal product or service		Number, street, and room or suite no. If a P.O. box, see the instructions.	**E** Date business started
C Business code number		City or town, state, and ZIP code	**F** Total assets (see the instructions) $

G Check applicable boxes: **(1)** ☐ Initial return **(2)** ☐ Final return **(3)** ☐ Name change **(4)** ☐ Address change **(5)** ☐ Amended return
(6) ☐ Technical termination - also check (1) or (2)

H Check accounting method: **(1)** ☐ Cash **(2)** ☐ Accrual **(3)** ☐ Other (specify) ▶ _____

I Number of Schedules K-1. Attach one for each person who was a partner at any time during the tax year ▶ _____

J Check if Schedules C and M-3 are attached . . . . . ☐

Caution. *Include **only** trade or business income and expenses on lines 1a through 22 below. See the instructions for more information.*

Income	**1a**	Merchant card and third-party payments (including amounts reported on Form(s) 1099-K). For 2011, enter -0-	**1a**			
	b	Gross receipts or sales not reported on line 1a (see instructions)	**1b**			
	c	Total. Add lines 1a and 1b	**1c**			
	d	Returns and allowances plus any other adjustments to line 1a (see instructions)	**1d**			
	e	Subtract line 1d from line 1c	**1e**			
	2	Cost of goods sold (attach Form 1125-A)	**2**			
	3	Gross profit. Subtract line 2 from line 1e			**3**	
	4	Ordinary income (loss) from other partnerships, estates, and trusts (attach statement)			**4**	
	5	Net farm profit (loss) (attach Schedule F (Form 1040))			**5**	
	6	Net gain (loss) from Form 4797, Part II, line 17 (attach Form 4797)			**6**	
	7	Other income (loss) (attach statement)			**7**	
	8	**Total income (loss).** Combine lines 3 through 7			**8**	

...which lists allocations only by partner type...

Form 1065 (2011) Page 5

Analysis of Net Income (Loss)

1	Net income (loss). Combine Schedule K, lines 1 through 11. From the result, subtract the sum of Schedule K, lines 12 through 13d, and 16l					1	
2	Analysis by partner type:	(i) Corporate	(ii) Individual (active)	(iii) Individual (passive)	(iv) Partnership	(v) Exempt organization	(vi) Nominee/Other
a	General partners						
b	Limited partners						

...but are reflected in K-1s (issued *per partner*)

Schedule K-1 (Form 1065)

Department of the Treasury
Internal Revenue Service

2011

For calendar year 2011, or tax year beginning _____, 2011
ending _____, 20____

Partner's Share of Income, Deductions, Credits, etc. ▶ See back of form and separate instructions.

☐ Final K-1 ☐ Amended K-1 OMB No. 1545-0099

Part I Information About the Partnership

A Partnership's employer identification number

B Partnership's name, address, city, state, and ZIP code

C IRS Center where partnership filed return

D ☐ Check if this is a publicly traded partnership (PTP)

Part II Information About the Partner

E Partner's identifying number

F Partner's name, address, city, state, and ZIP code

Part III Partner's Share of Current Year Income, Deductions, Credits, and Other Items

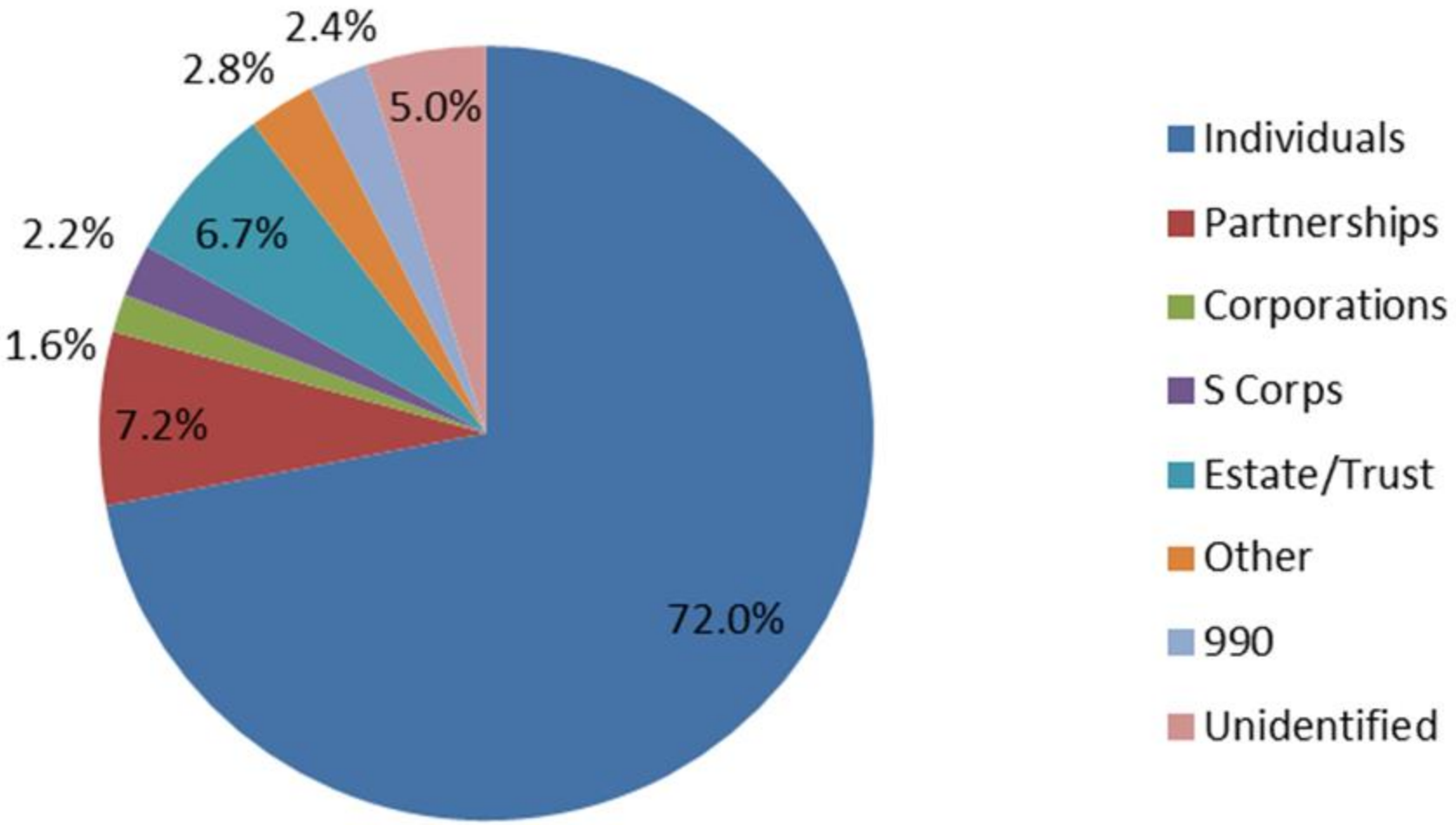
1	Ordinary business income (loss)	15	Credits
2	Net rental real estate income (loss)		
3	Other net rental income (loss)	16	Foreign transactions
4	Guaranteed payments		
5	Interest income		
6a	Ordinary dividends		
6b	Qualified dividends		
7	Royalties		
8	Net short-term capital gain (loss)		
9a	Net long-term capital gain (loss)	17	Alternative minimum tax (AMT) items
9b	Collectibles (28%) gain (loss)		
9c	Unrecaptured section 1250 gain		
10	Net section 1231 gain (loss)	18	Tax-exempt income and nondeductible expenses
11	Other income (loss)		

Linking partnerships to partners

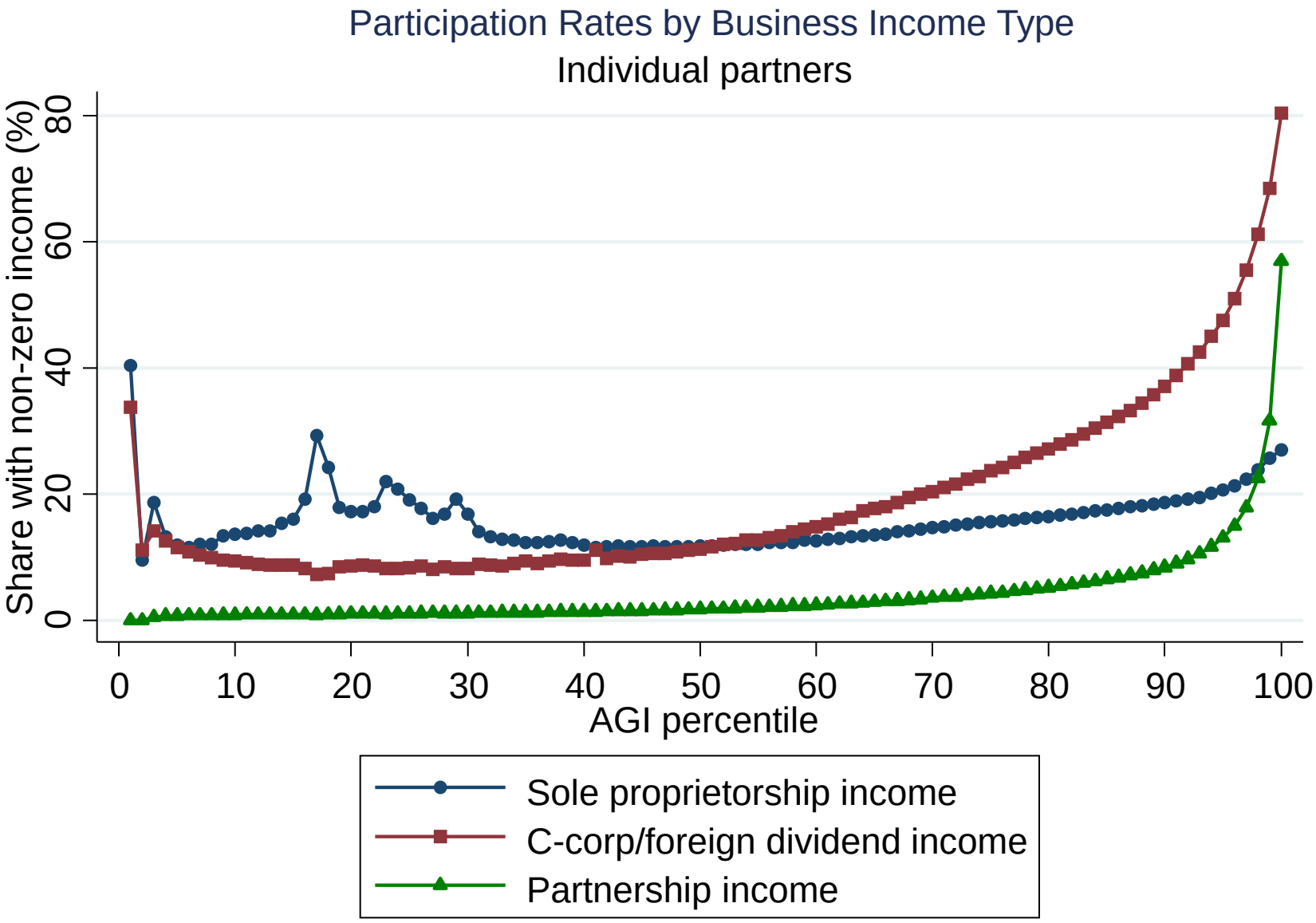
- Data challenges: Owners can be one of many entity types, and different kinds of returns are processed by different systems

- Our procedure: link K-1's to partnership returns by merging on the Document Locator Number (linking **25m** K-1's to **3m** partnerships)
    - High coverage: >96% of business income

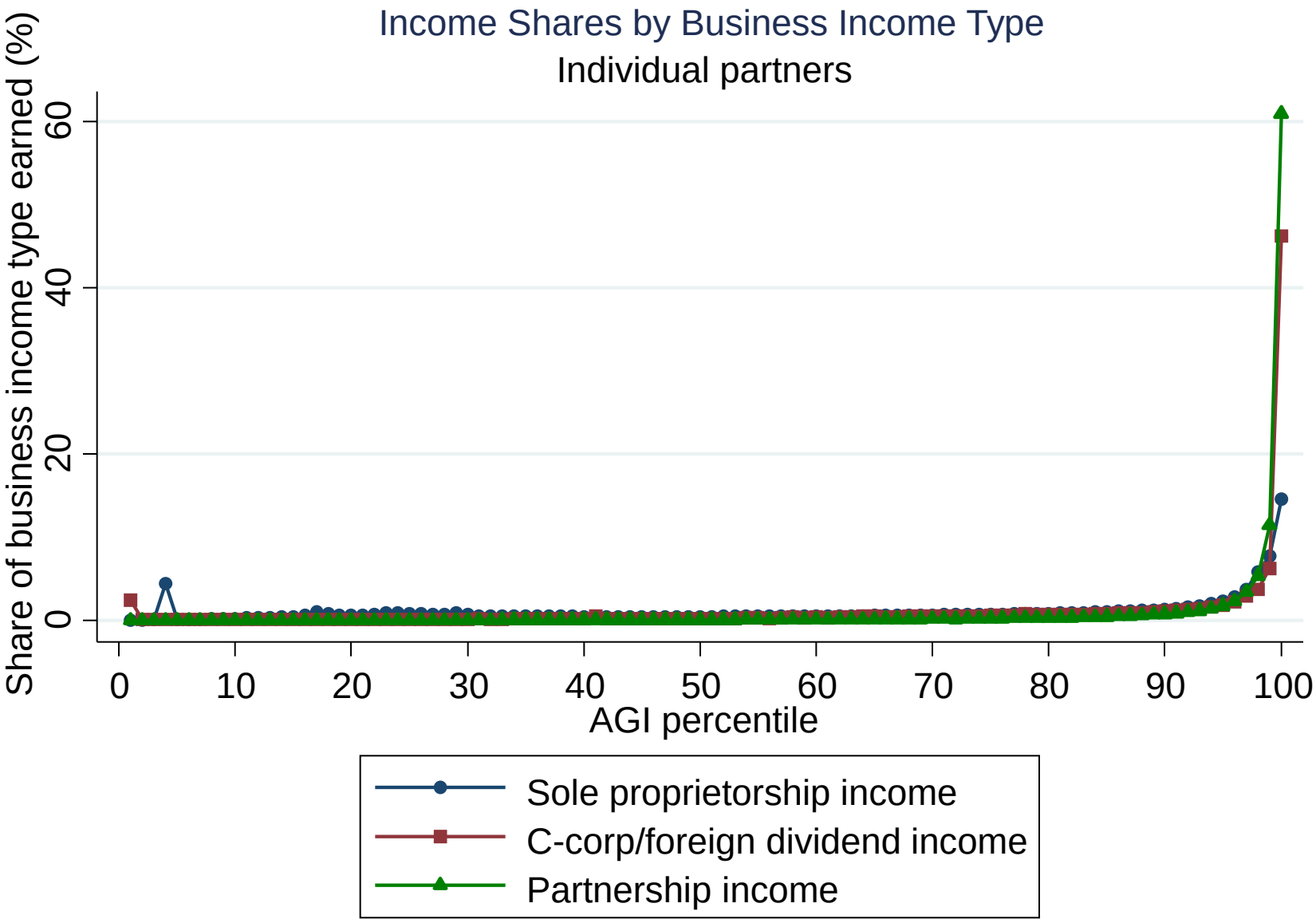
Part 2: Who owns partnerships



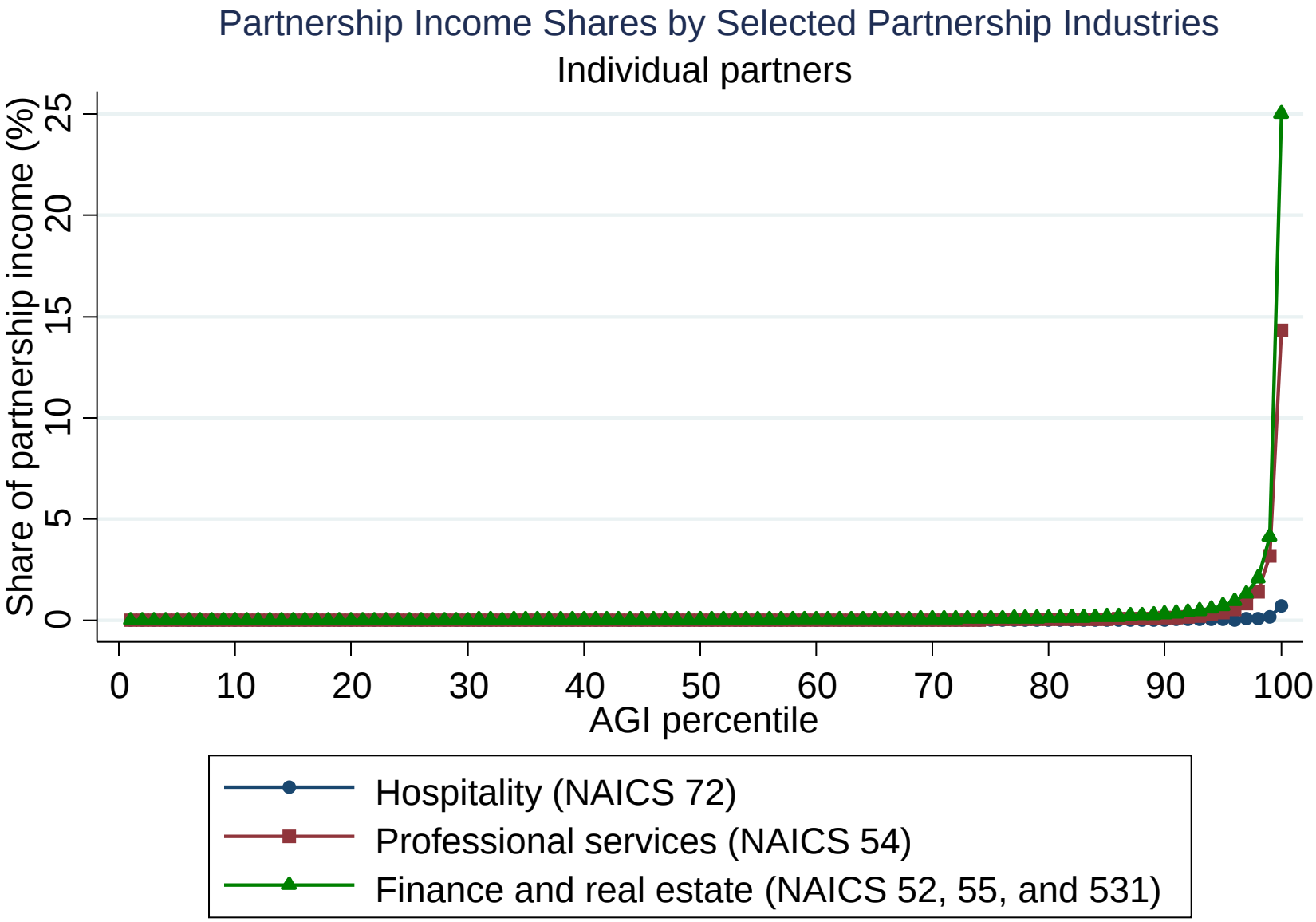
Partnership participation is very concentrated



Partnership income is exceptionally concentrated



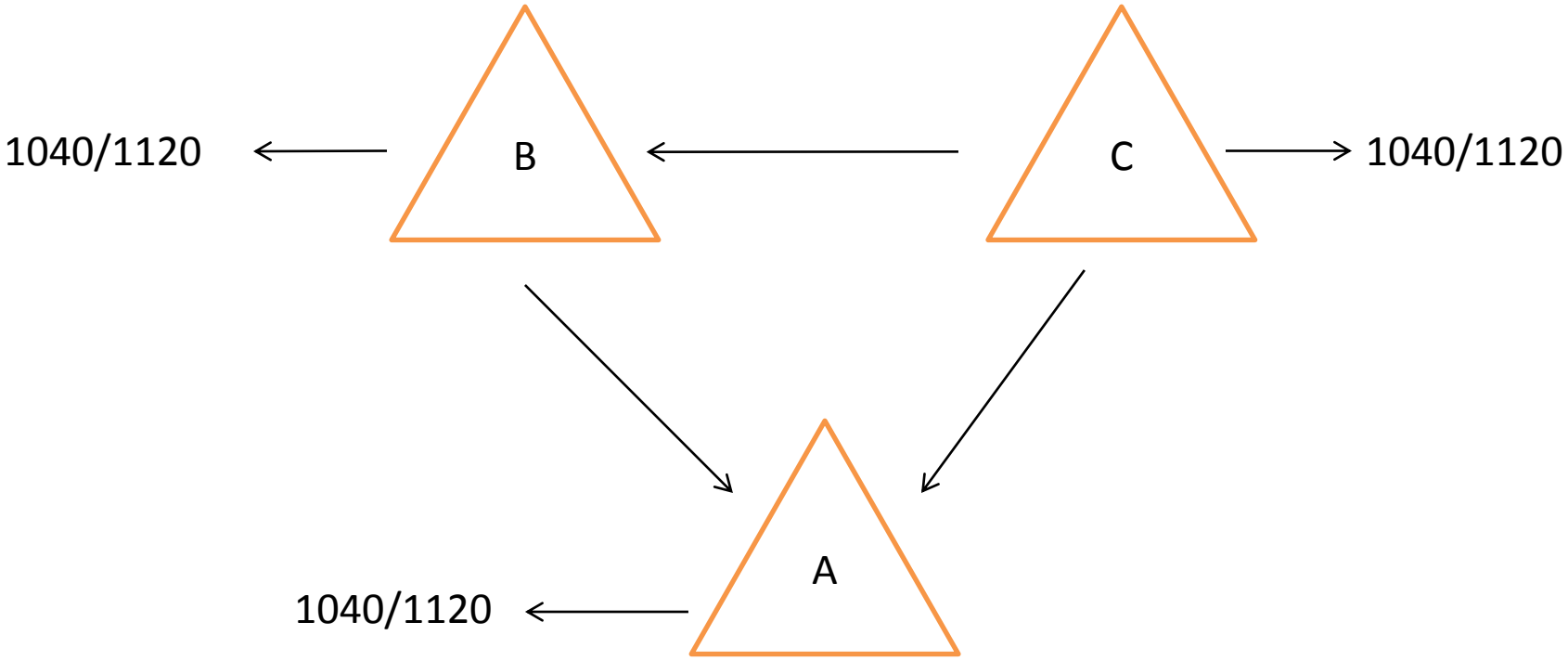
High-income partners own finance / prof. services



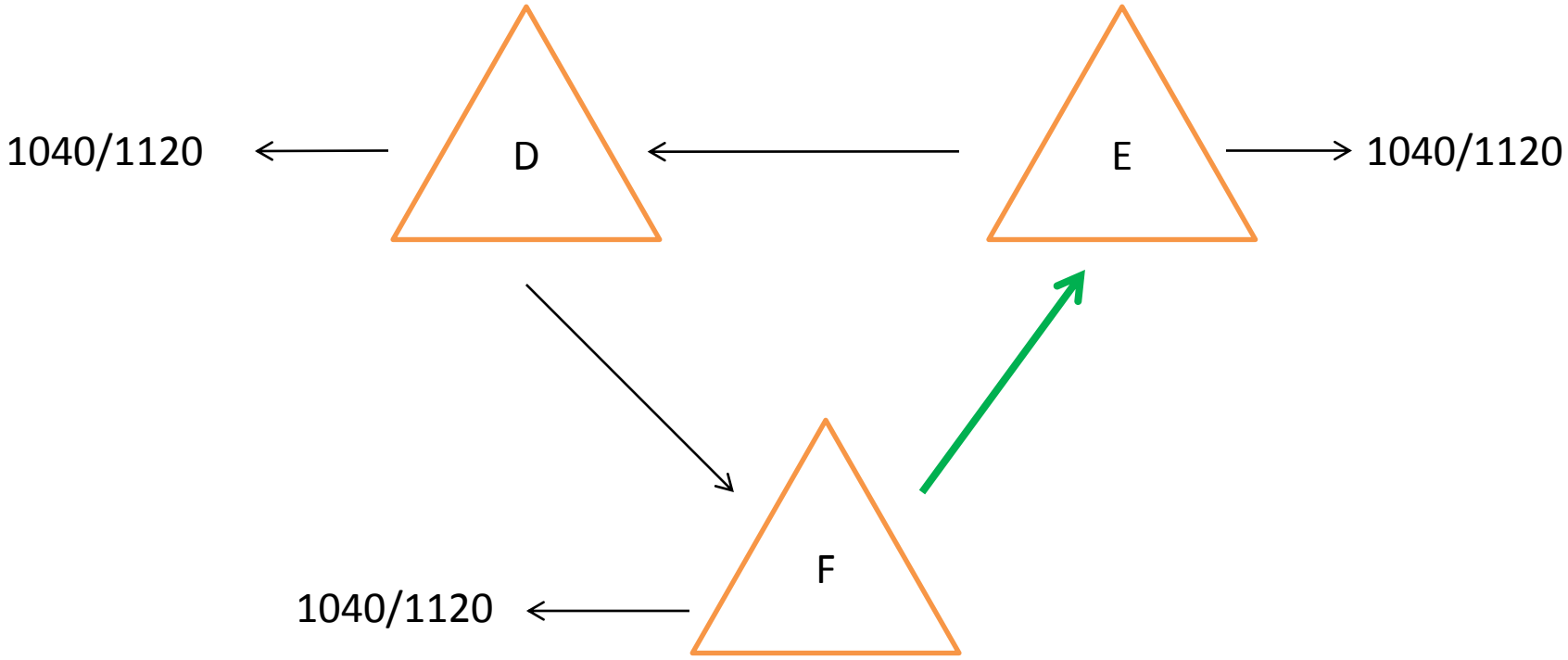
Part 3: Estimating partnership tax rates

- Assign tax rates to each partner
  - For tiered partnerships this requires tracking flows through the tiers to a final owner
  - For S-corporation partners this requires assigning a tax rate to each shareholder of the S-corporation
- Aggregate circular tiered partnerships into single partnerships
- Aggregate these partner tax rates to the partnership level

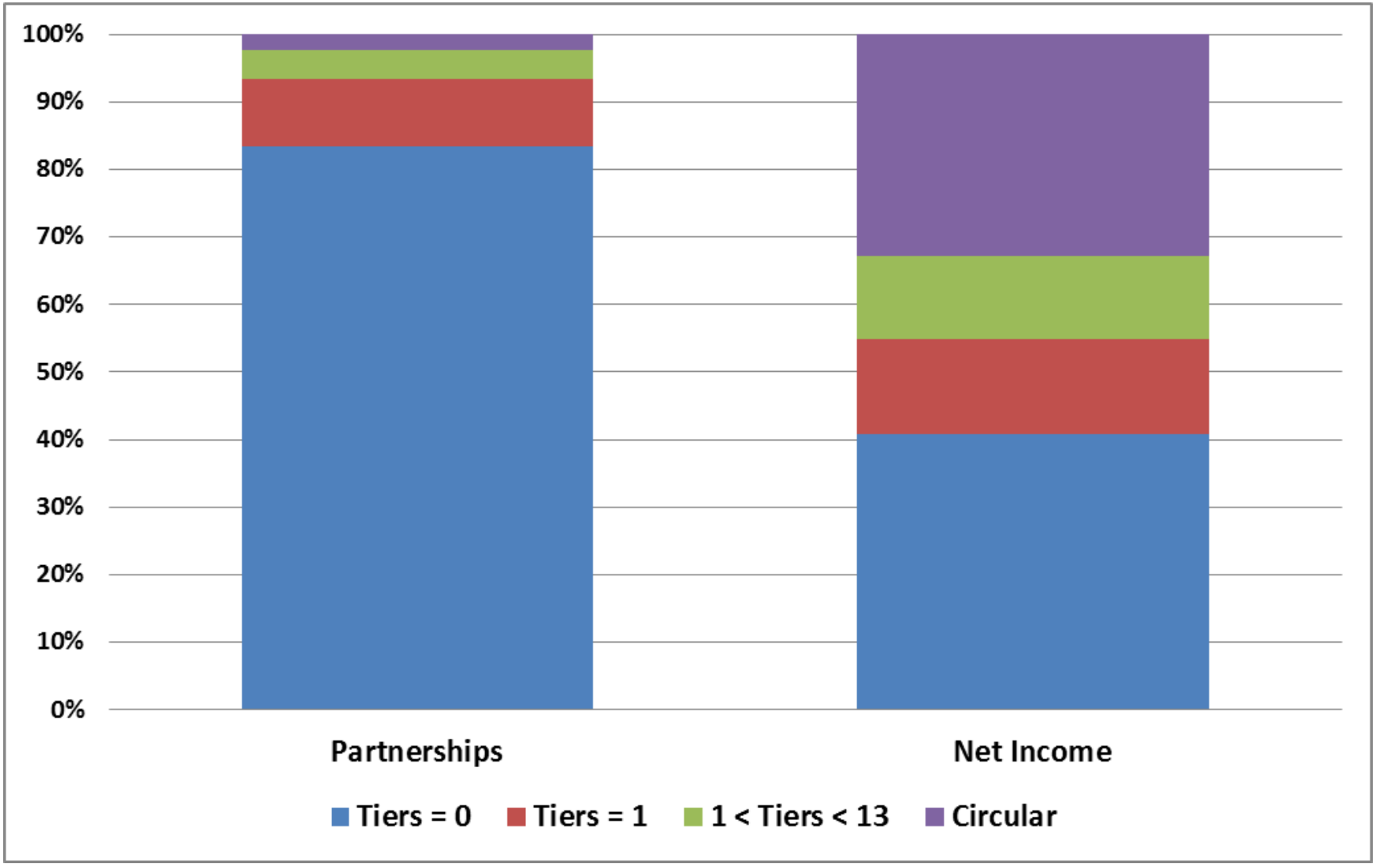
Case 1: Linear tiers



Case 2: Circular tiers (treat as single partnership)



Tiered partnerships



Assign tax rates to partners

- Each partner's income tax liability is recalculated w/o the allocated amount of each income type appearing on the K-1 from a given partnership
- OTA has CDW-based tax calculators for 1040, 1041, 1120-C and 1120-S
- Those without a calculator:
  - 1120-F: 35%
  - 1120-L: 35%
  - 1120-PC: 35%
  - 1120-REIT: 35%
  - 1120-RIC: 22%
  - 990: 35% (UBIT) or 0%
  - Unidentified: 35%

Aggregate partner tax rates to a partnership

- Define: Tax rate on income type i to partner k from partnership p is:

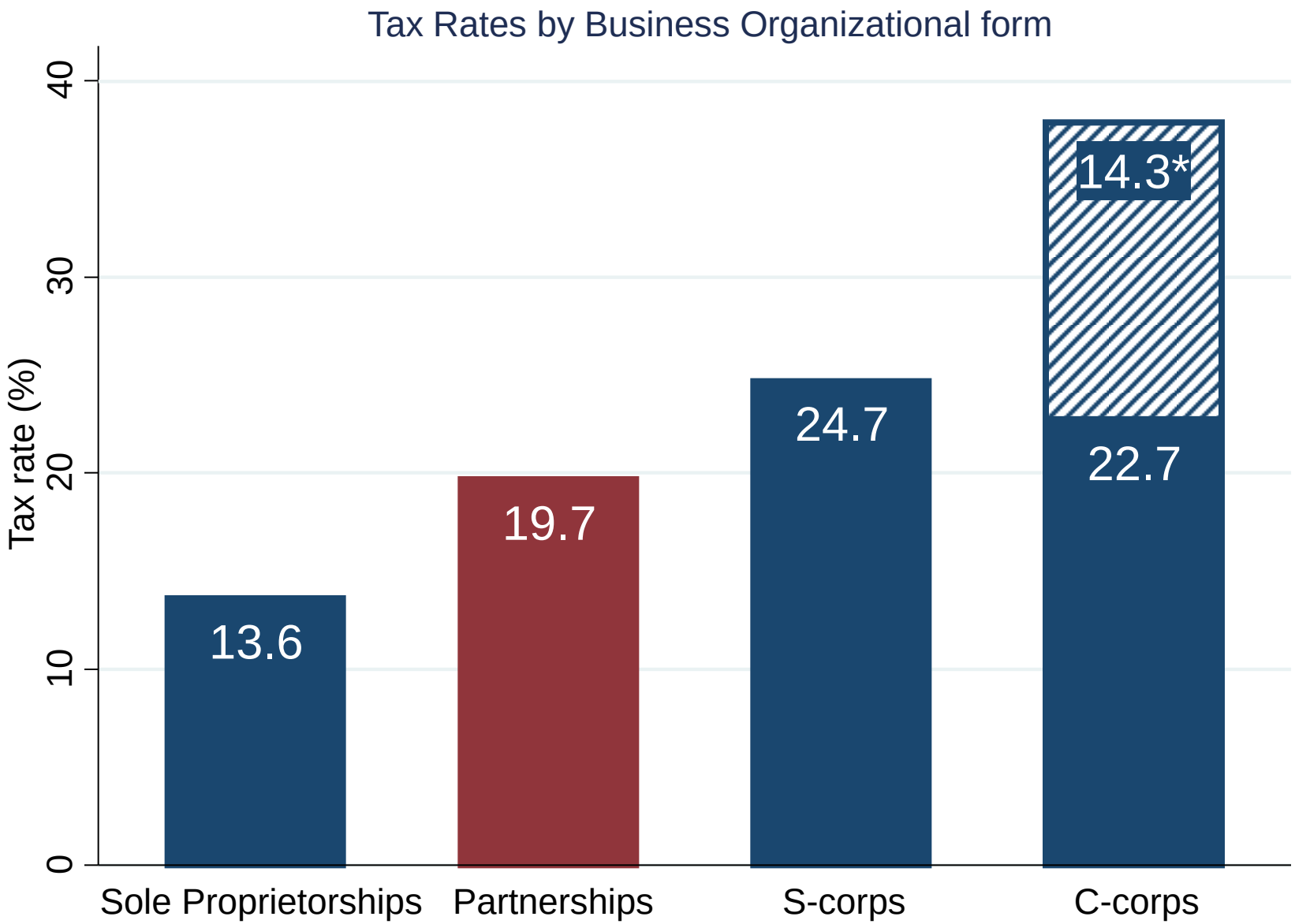
$$T_{i,k,p} = \frac{TAX_{i,k,p}}{D_{i,k,p}}$$

- These rates can be aggregated up for a partnership tax rate:

$$T_p = \frac{1}{D_p}\left[\sum_{i=1}^{I}\sum_{k=1}^{K} D_{i,k,p} * T_{i,k,p}\right]$$

- Note: Definitions matter
  - Tax rates technically unbounded but vast majority in [0%,35%]
  - Weighting by actual amount vs. absolute value matters at top

Results: Entity income tax rates by entity form



*Source: Based on 18.5% average U.S. dividend rate estimated in Poterba (2004)

Summary

1. Who owns them?: Partnership income accrues to high-earners even more disproportionately than C-corporate income (60% to the top 1%)

2. How much tax do they pay?: We estimate a partnership tax rate of 19.7%, lower than the C-corporate and lower for the largest partnerships

Suggestions for SOI products

- Future OTA partnership work will rely on links to ultimate owners

- The two key uses of our linked data could hopefully be incorporated into SOI's partnership study file (and potentially other business study files):

  1. Adding an ownership file, similar to S-corporation study
  2. Including all K-1 fields in ownership file

1: Adding an ownership file, similar to S-corp. study

- S-corporation study file: Comes with information on S-corporation owners
- Partnerships: Much harder because of partnership tiers

- Suggestion: Provide one "entity" file and one "ownership" file
  - Entity file: Current partnership study file
  - Ownership file: Direct and indirect owners of entity file partnerships
    - Sample: Owners of the partnerships in the entity file and owners at least three tiers down
    - Rows: One row per owner
    - Columns: K-1 fields and indicator of partnership tier

2: Including all K-1 fields in ownership file

Schedule K-1 (Form 1065)

Department of the Treasury
Internal Revenue Service

2011

For calendar year 2011, or tax year beginning _____, 2011
ending _____, 20____

Partner's Share of Income, Deductions, Credits, etc. ▶ See back of form and separate instructions.

☐ Final K-1 ☐ Amended K-1 OMB No. 1545-0099

Part I Information About the Partnership

A Partnership's employer identification number

B Partnership's name, address, city, state, and ZIP code

C IRS Center where partnership filed return

D ☐ Check if this is a publicly traded partnership (PTP)

Part II Information About the Partner

E Partner's identifying number

F Partner's name, address, city, state, and ZIP code

Part III Partner's Share of Current Year Income, Deductions, Credits, and Other Items

1	Ordinary business income (loss)	15	Credits
2	Net rental real estate income (loss)		
3	Other net rental income (loss)	16	Foreign transactions
4	Guaranteed payments		
5	Interest income		
6a	Ordinary dividends		
6b	Qualified dividends		
7	Royalties		
8	Net short-term capital gain (loss)		
9a	Net long-term capital gain (loss)	17	Alternative minimum tax (AMT) items
9b	Collectibles (28%) gain (loss)		
9c	Unrecaptured section 1250 gain		
10	Net section 1231 gain (loss)	18	Tax-exempt income and nondeductible expenses
11	Other income (loss)		

What is SOI's Value Added?

Lessons from the Non-Filer Project

June 5, 2015
Jim Nunns
SOI Consultants Panel Meeting

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

SOI's Value Added: Context

- CDW creates vast menu for new value added
  - New products, improved population data
- SOI eager to enhance products and services
- But, SOI faces severe resource constraints
  - Tight budgets into foreseeable future
  - Days of "easy" productivity gains over
  - Human capital hard to develop & maintain
  - Ongoing demand for current SOI products
- Tradeoffs and careful study design required

SOI's Value Added: Origins of Non-Filer Project

- Non-Filer Project shows how CDW adds value

Origins

- Most of population represented on tax returns
  - "Filers" may be taxpayers or dependents
- Information returns cover most "Non-Filers"
- Early SOI studies on non-filer population
- Recent micro-data studies
  - JCT and OTA for microsimulation models
  - RAS for National Research Program (NRP)

SOI's Value Added: Elements of Non-Filer Project

- Collaborative effort between SOI & customers
- Based on sample of information returns
- Clean data; add SSA data; flag nonresidents
- Match to CDW to identify filers
- Map information into "return" for each person
- Final products:
  - Micro data files for INSOLE and PUF
  - Tabulations covering filers and non-filers

SOI's Value Added: Lessons from Non-Filer Project

- Creates new low-cost, high-value products
- Draws on comparative advantages
  - "Experimental" work by JCT, OTA, RAS
  - "Gold standard" production by SOI
- Builds on existing SOI products
- CDW essential to success
- Methodology well documented
- Straightforward extensions to population files
  - Improve data on CDW and Data Bank
  - Adds geographic and longitudinal depth

THANK YOU

For more information please contact:

Jim Nunns
jnunns@urban.org

Visit us at:
www.taxpolicycenter.org

TPC TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

AAPOR Report on Big Data

AAPOR Big Data Task Force

February 12, 2015

Task Force Members:

***Lilli Japec**, Co-Chair, Statistics Sweden*

***Frauke Kreuter**, Co-Chair, JPSM at the U. of Maryland, U. of Mannheim & IAB*

***Marcus Berg**, Stockholm University*

***Paul Biemer**, RTI International*

***Paul Decker**, Mathematica Policy Research*

***Cliff Lampe**, School of Information at the University of Michigan*

***Julia Lane**, American Institutes for Research*

***Cathy O'Neil**, Johnson Research Labs*

***Abe Usher**, HumanGeo Group*

Acknowledgement: We are grateful for comments, feedback and editorial help from Eran Ben-Porath, Jason McMillan, and the AAPOR council members.

The report has four objectives:

1. to educate the AAPOR membership about Big Data (Section 3)
2. to describe the Big Data potential (Section 4 and Section 7)
3. to describe the Big Data challenges (Section 5 and 6)
4. to discuss possible solutions and research needs (Section 8)

Big Data

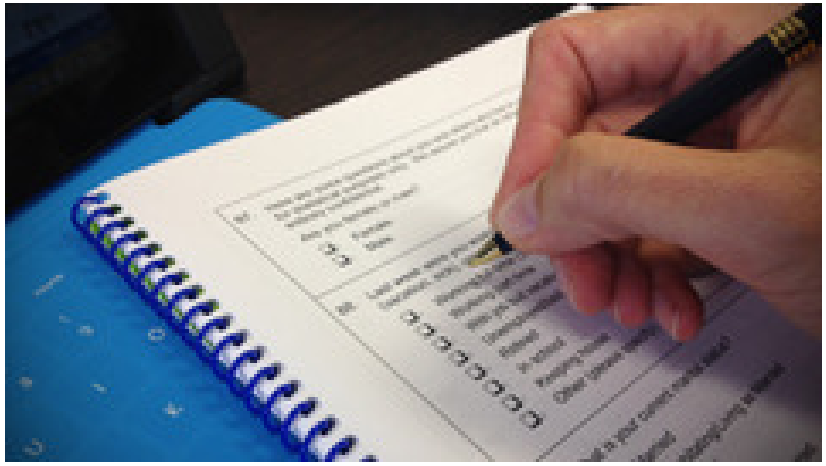
AAPOR Task Force

Source: Frauke Kreuter

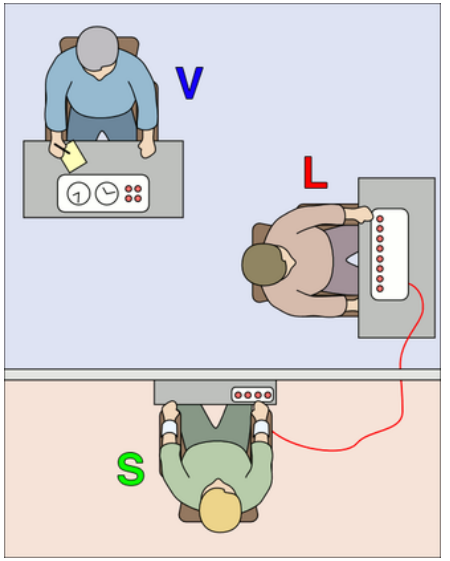
until recently

three main data sources

Survey Data



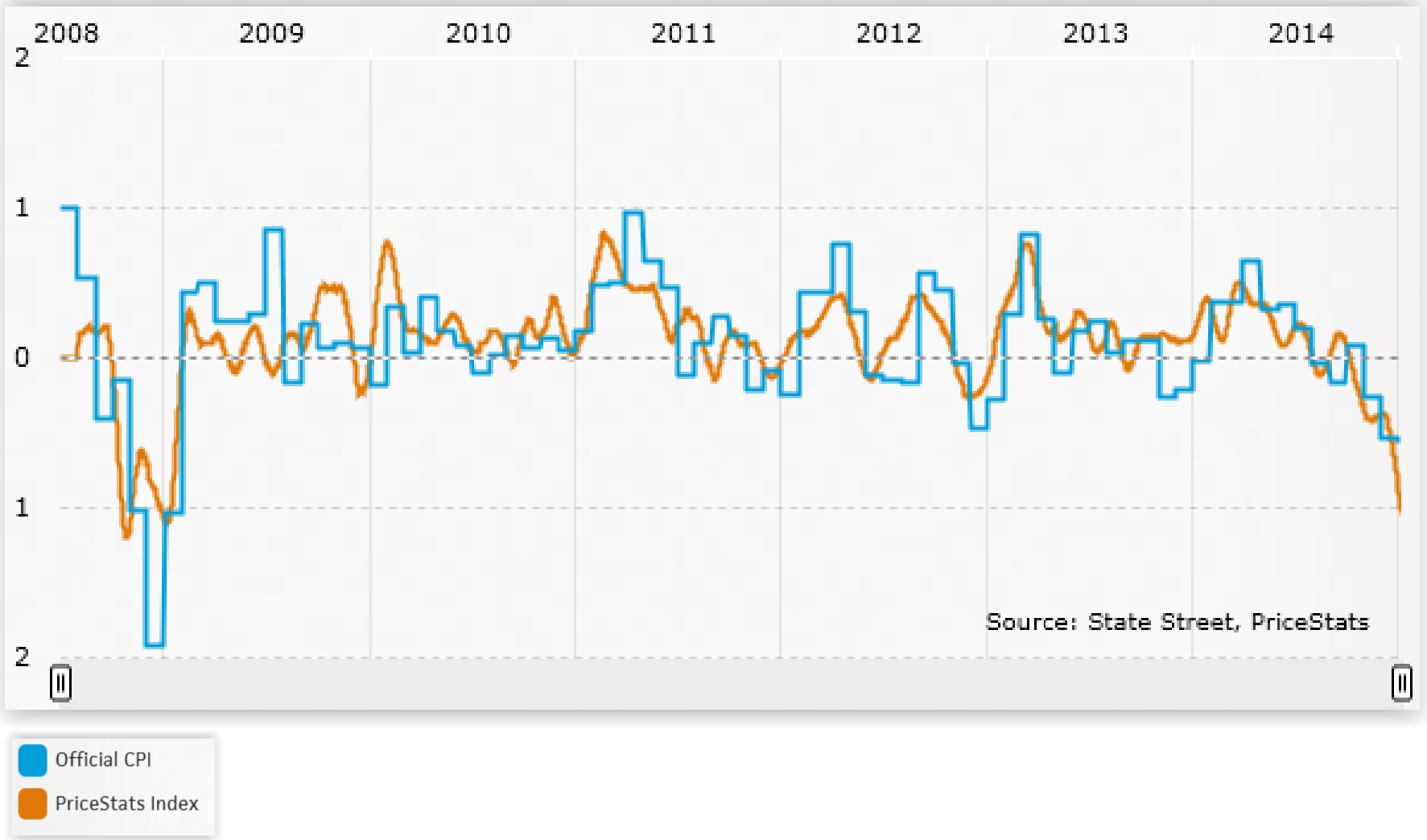
Experiments



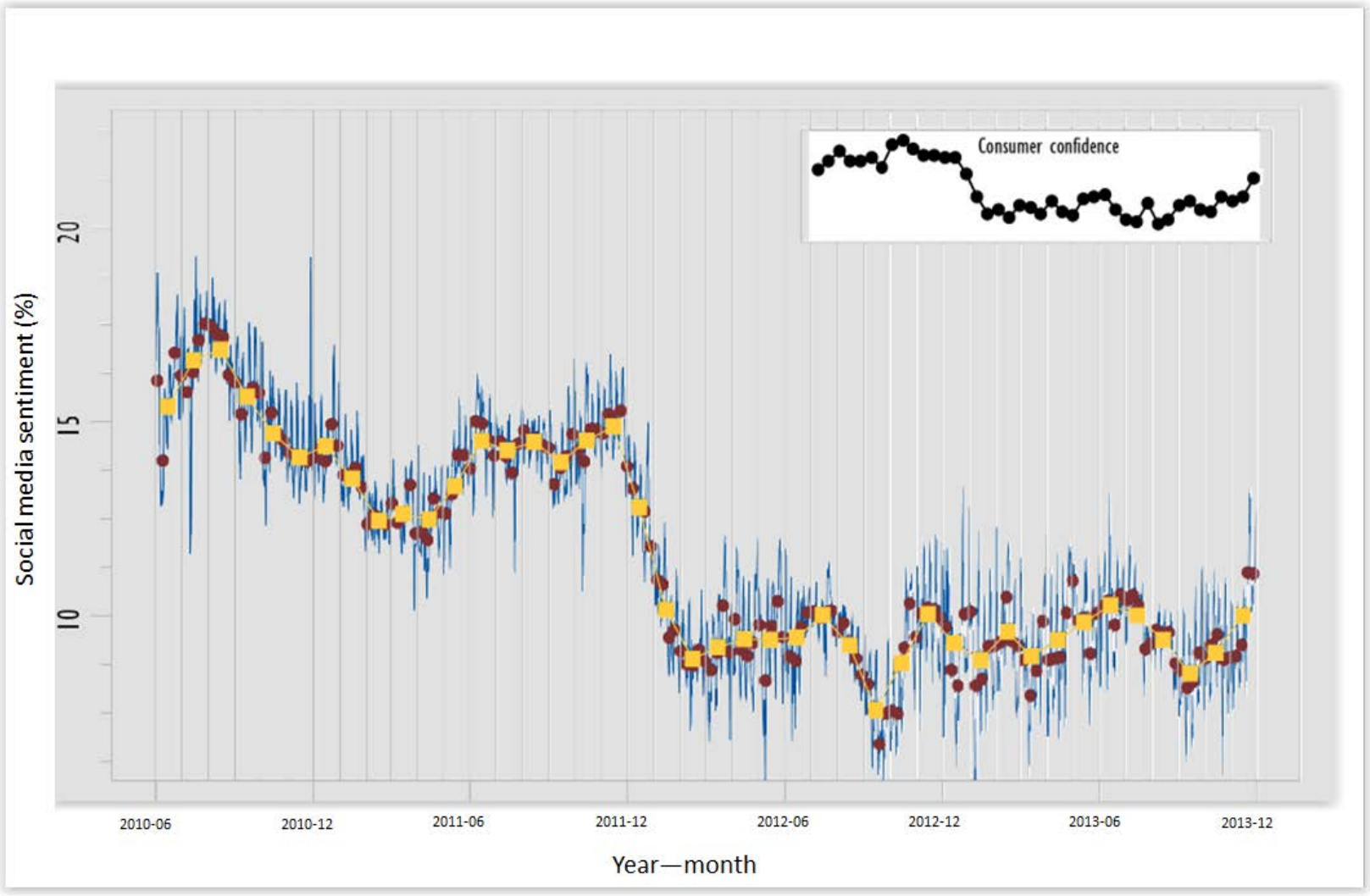
Administrative Data

Source: Frauke Kreuter

now

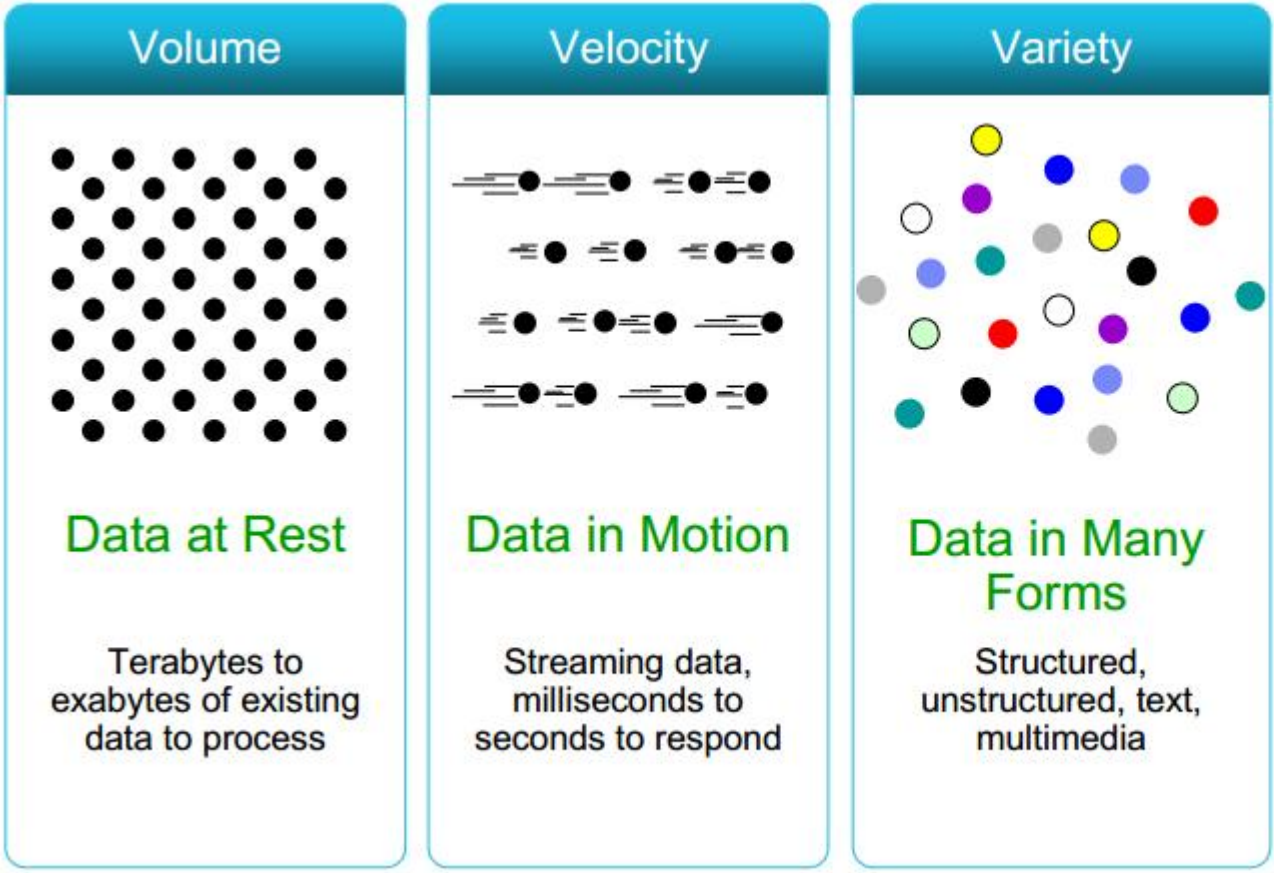


US Aggregated Inflation Series, Monthly Rate, PriceStats Index vs. Official CPI. Accessed January 18, 2015 from the PriceStats website.



Social media sentiment (daily, weekly and monthly) in the Netherlands, June 2010 - November 2013. The development of consumer confidence for the same period is shown in the insert (Daas and Puts 2014).

Big Data



http://www.rosebt.com/blog/data-veracity

Hope that found/organic data

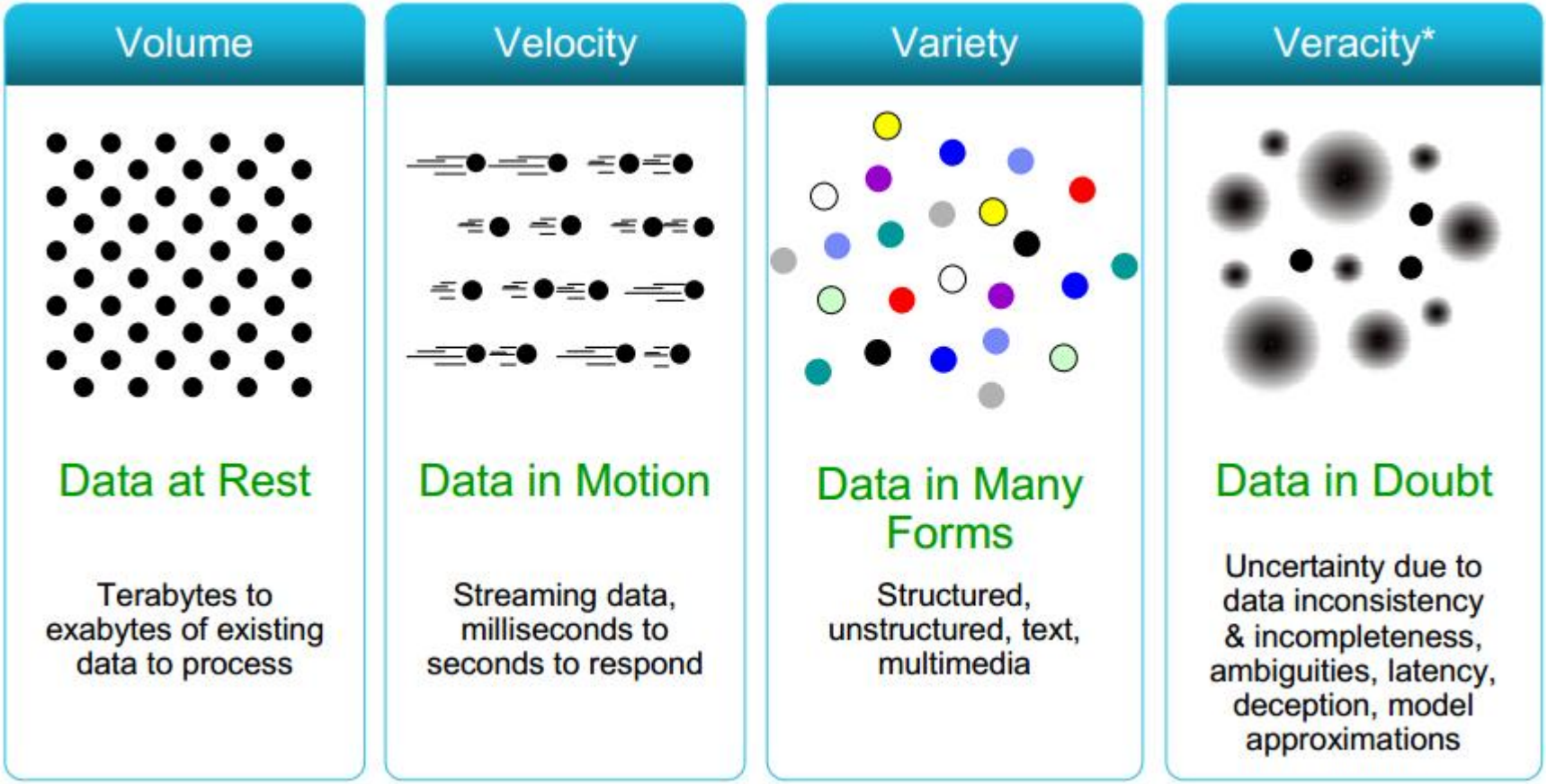
Can replace or augment expensive data collections

More (= better) data for decision making

Information available in (nearly) real time

Source: Frauke Kreuter

But (at least) one more V



http://www.rosebt.com/blog/data-veracity

fkreuter@umd.edu

Thank You!

CHANGE IN PARADIGM AND RISKS INVOLVED

Julia Lane
New York University
American Institutes for Research
University of Strasbourg

Big Data definition

- “Big Data” is an imprecise description of a rich and complicated set of characteristics, practices, techniques, ethics, and outcomes all associated with data. (AAPOR)
- No canonical definition
- By characteristics: Volume Velocity Variety (and Variability and Veracity)
- By source: found vs. made
- By use: professionals vs. citizen science
- By reach: datafication
- By paradigm: Fourth paradigm

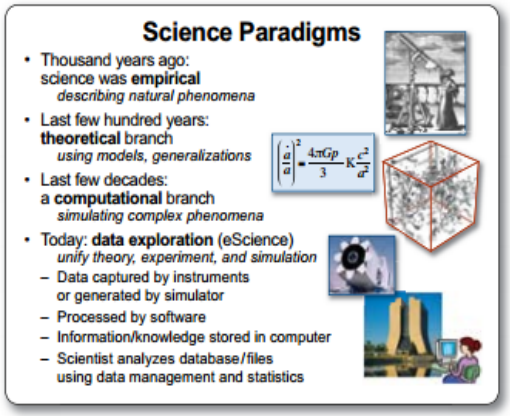


FIGURE 1

Source: Julia Lane

Motivation

- New business model
  - Federal agencies no longer major players
- New analytical model
  - Outliers
  - Finegrained analysis
  - New units of analysis
- New sets of skills
  - Computer scientists
  - Citizen scientists
- Different cost structure

Big Data

Jim Gray's paradigm

- **Observational Science**
  - Scientist gathers data by direct observation
  - Scientist analyzes data
- **Analytical Science**
  - Scientist builds analytical model
  - Makes predictions.
- **Computational Science**
  - Simulate analytical model
  - Validate model and makes predictions
- **Data Exploration Science**
  - **Data-driven science**
    Data captured by instruments or from the web, or data generated by simulation
  - Information extraction
  - Processed by software
  - Placed in a database / files
  - Scientist(s)/scholar(s) analyze(s) database / files
  - *Access crucial*

Training to Climb an Everest of Digital Data

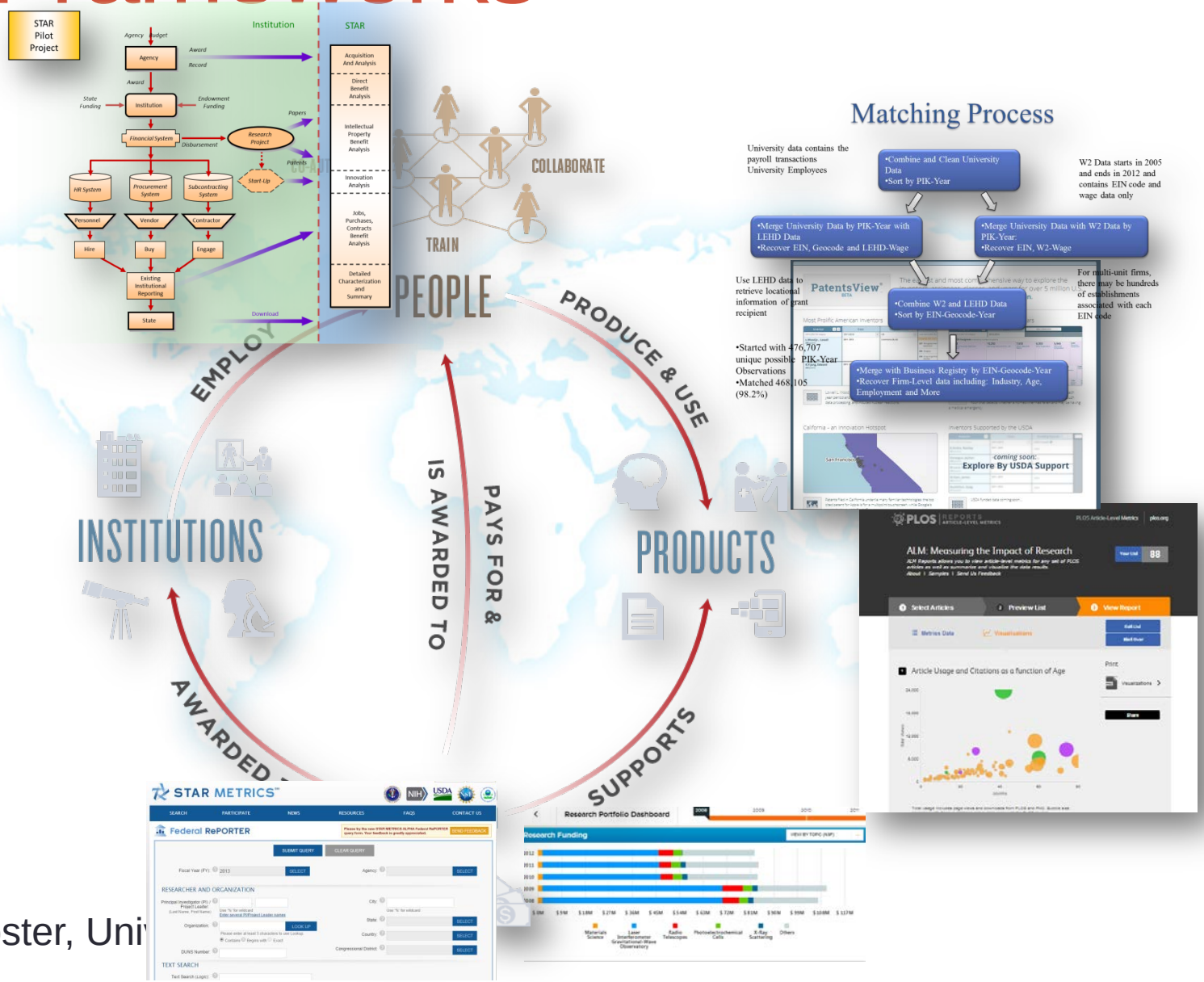
By ASHLEE VANCE
Published: October 11, 2009

MOUNTAIN VIEW, Calif. — It is a rare criticism of elite American university students that they do not think big enough. But that is exactly the complaint from some of the largest technology companies and the federal government.

Source: Lee Giles

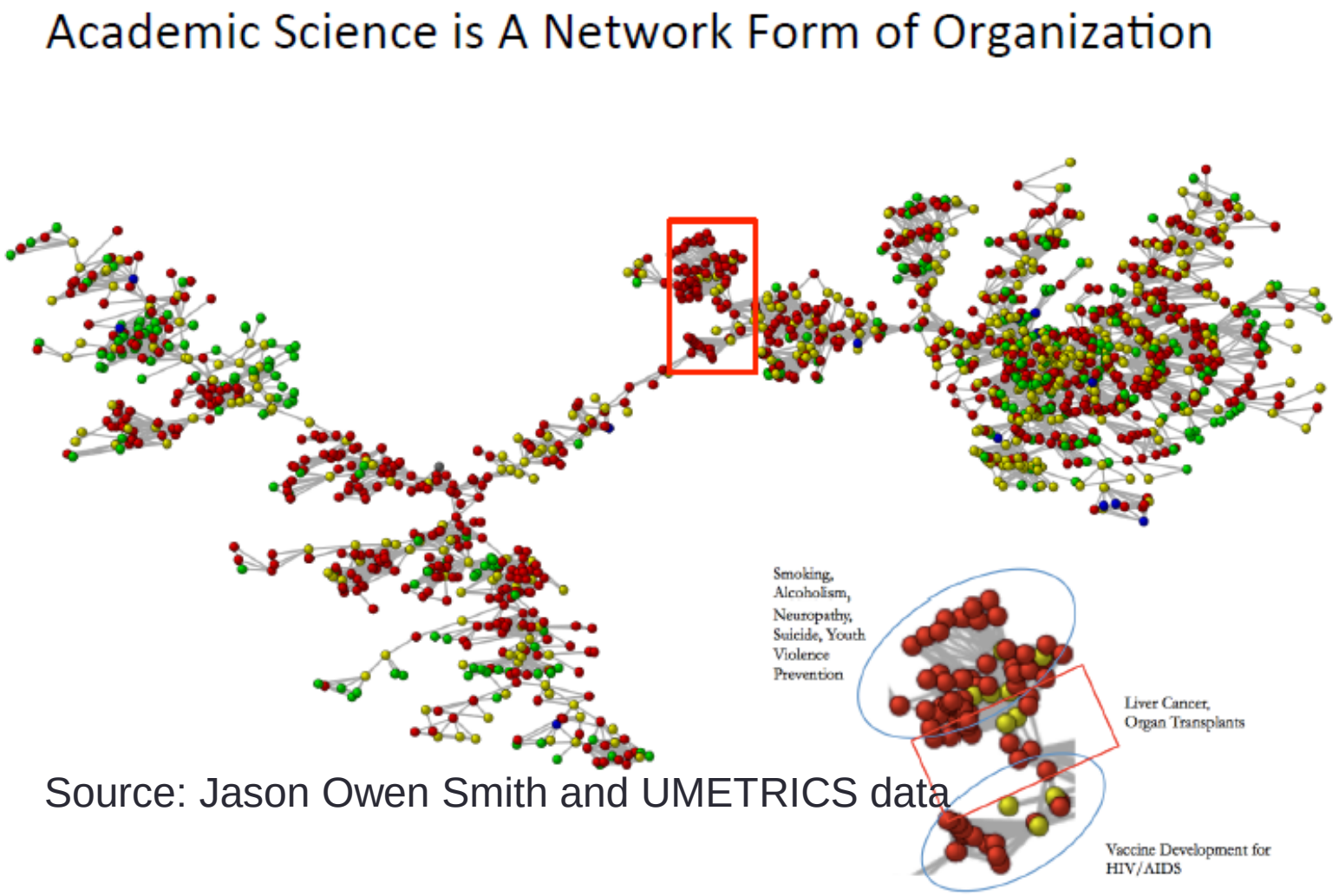
Source: Julia Lane

New Frameworks



Source: Ian Foster, Uni

New kinds of analysis



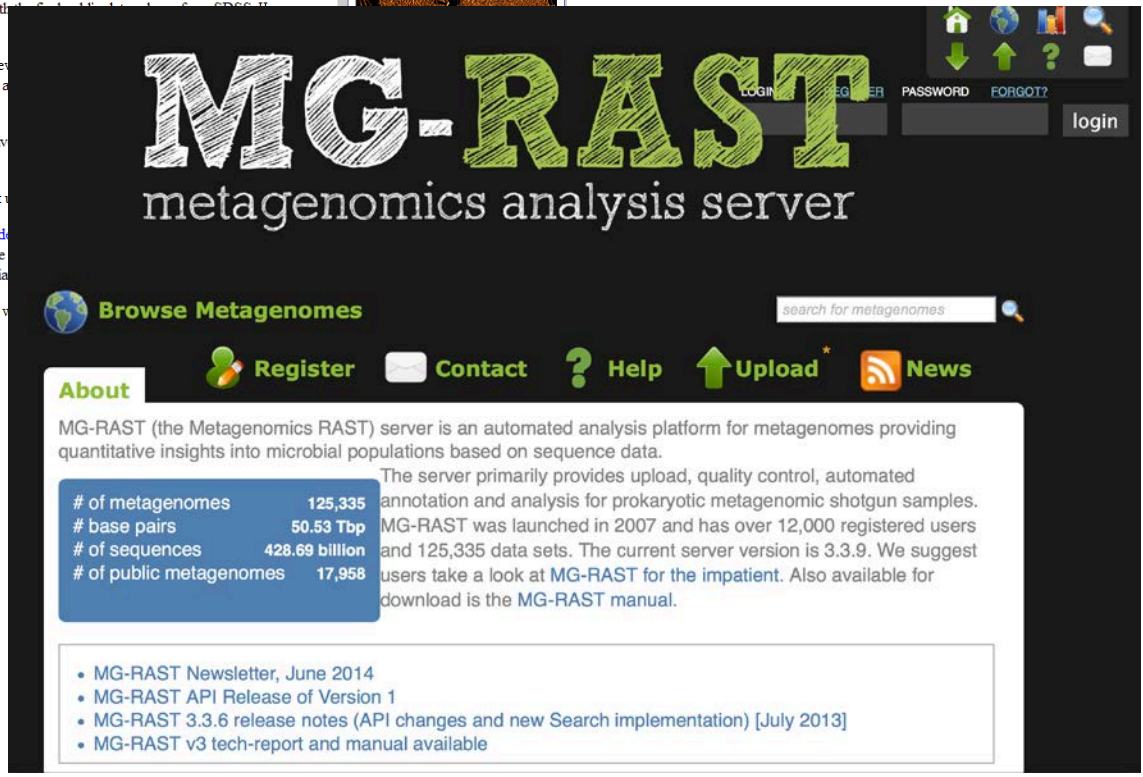
Source: Jason Owen Smith and UMETRICS data

Access for Research

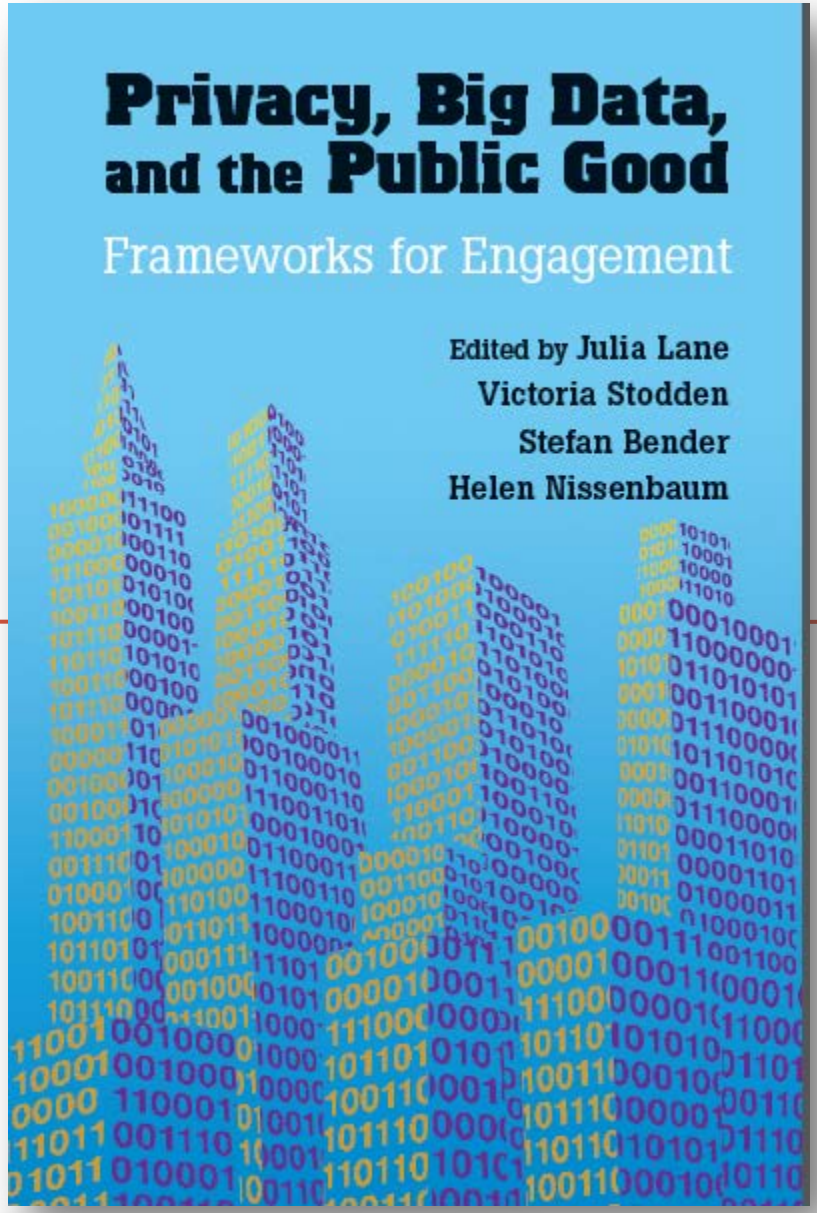
Science measurement

Source: Julia Lane

Value in other fields



Source: Julia Lane



Source: Julia Lane

Core Questions

- What is the legal framework?
- What is the practical framework?
- What is the statistical framework?

Source: Julia Lane

Le

ResearchGate

Sign up | Log

Ron S. Jarmin

Ph.D.

Assistant Director

U.S. Census Bureau · Research and Methodology

17.59

Follow

Student wrongly tied to Boston bombings found dead

Doug Stanglin, USA TODAY 9:07 p.m. EDT April 25, 2013

Reddit apologized for the 'dangerous speculation' on the site that pointed fingers at the student.

(Photo: AP)

41737 CONNECT | 2767 TWEET | 31 LINKEDIN | 376 COMMENT | EMAIL | MORE

A body pulled from the water off India Point Park in Rhode Island has been identified as the Brown University student mistakenly linked by amateur sleuths on a social media site to the Boston bombings.

The body of 22-year-old Sunil Tripathi was identified through dental records, Health Department spokeswoman Dara Chadwick said

USA NOW

UW student: SAE frat screamed 'you apes' at us

by Ron S. Jarmin

Join now to access them and millions of other **full-texts for free**.

Last name

Join for free

Source: Julia Lane

Informed consent (Nissenbaum)

Public lives of others

- "With friends like these…"
  - *Social networks: what friends reveal implicitly & explicitly*
- Tyranny of the Minority
  - *Inference from representative sample*
  - "multiple attributes can be inferred globally when as few as 20% of the users reveal their attribute information."

Mislove et al., "You Are Who You Know: Inferring User Profiles in Online Social Networks."

Source: Julia Lane

Statistical Framework

- Importance of valid inference
  - The role of statisti
- Inadequate [illegible]istical disclosure limitation
  - Diminished role of [illegible]gencies
  - Limitations of surv
- New analytical framework:
  - Mathematically rigorous theory of privacy
  - Measurement of privacy loss
  - Differential privacy

Source: Julia Lane

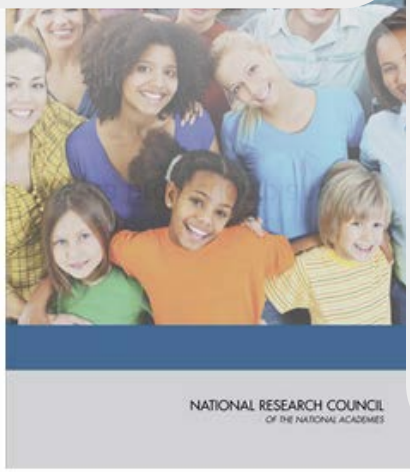
Some suggestions

Recommendations: Data Protection

Do not adopt HIPAA as the standard for data protection

Use an array of data protection approaches (Rec 5.1), such as:

- Plan with the concept of a portfolio approach considering safe people, safe projects, safe data, safe settings, and safe outputs
- Use a range of statistical methods to reduce disclosure risk
- Consult resources and data protection models, such as: university research data management service groups, individual IT/protection experts, and specialized institutions
- Use existing standards for data protection promulgated by the National Institute of Standards and Technology
- Develop a national center to define and certify information risk of different types of studies and corresponding data protection plans to minimize risks (Rec 5.2)



Recommendations for Research on Minimal Risk and Expedited Review

- Build evidence of risks in daily life and age-indexed routine medical, psychological, or educational examinations, tests, or procedures of th general population
- Develop appropriate algorithms for calculating risk from both the probability and magnitude of harm
- Encourage evidence for effective procedures for minimizing potential harms to no-more-than-minimal risk
- Study effects of social and behavioral research on research participar for evidence-based assessments of “known and foreseeable” risk

Source: Julia Lane

And a reminder of why

Source: Julia Lane

Comments and questions

- Julia.lane@nyu.edu

HUMANGEO

Turning information into insight

Skills Required to Integrate Big Data into Public Opinion Research

Abe Usher
Chief Technology Officer, HumanGeo

HUMANGEO

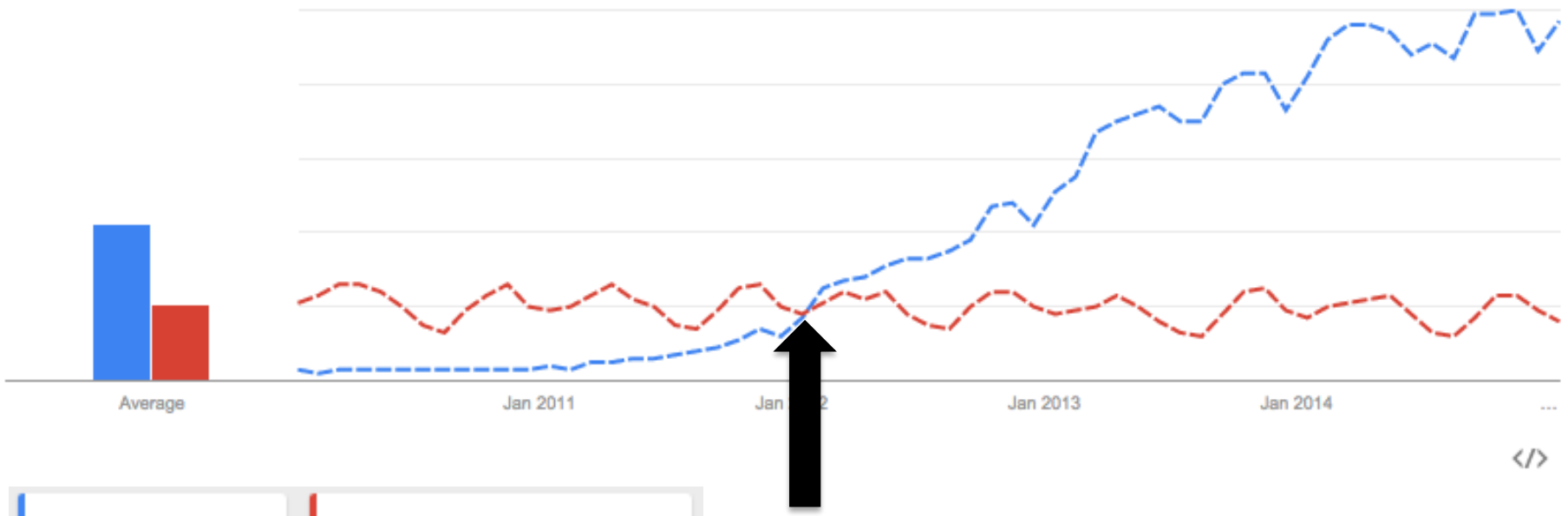
Outline

- Big data demystified
- Four layers of big data
- Skills required
- Easter eggs

Source: Abe Usher

HUMANGEO

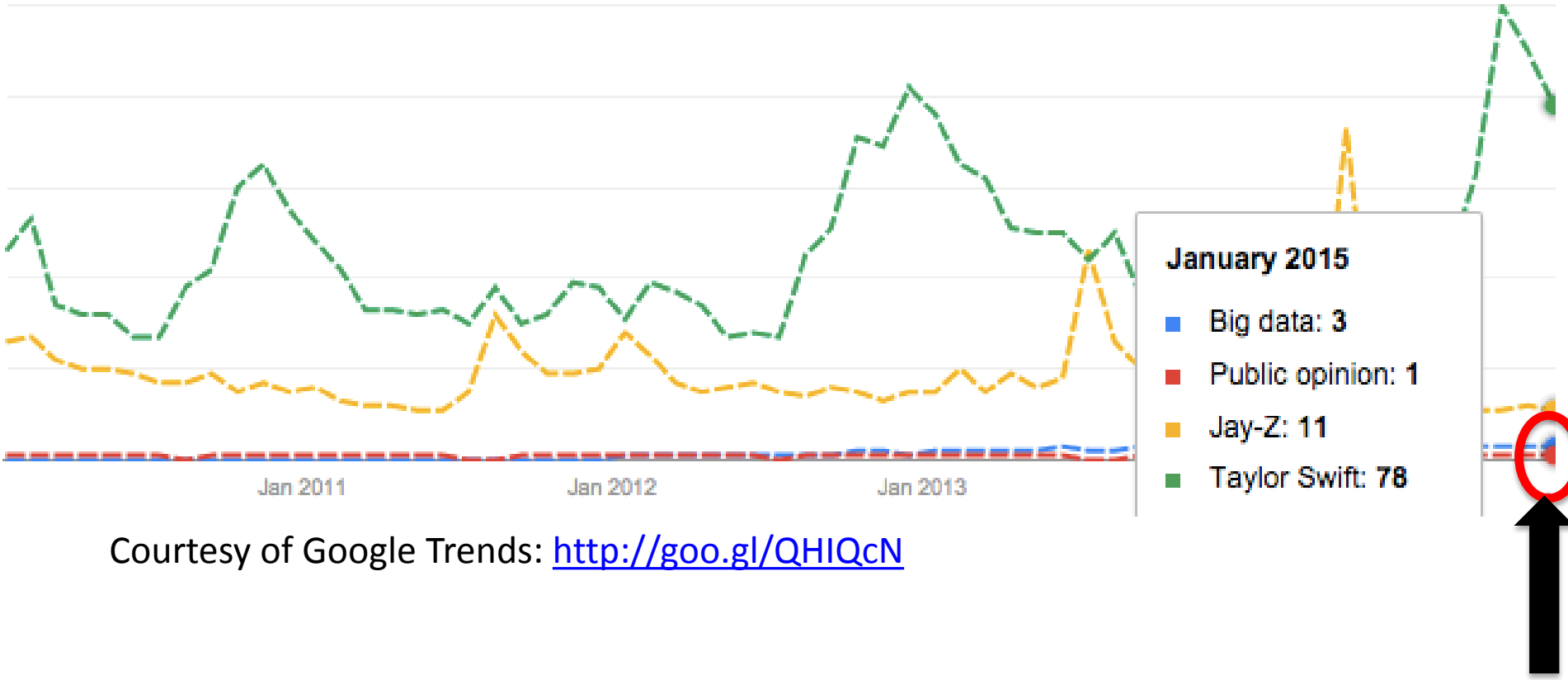
Big Data Today



Courtesy of Google Trends: http://goo.gl/4H8Ttd

HUMANGEO

Big Data & Public Opinion vs Pop Culture



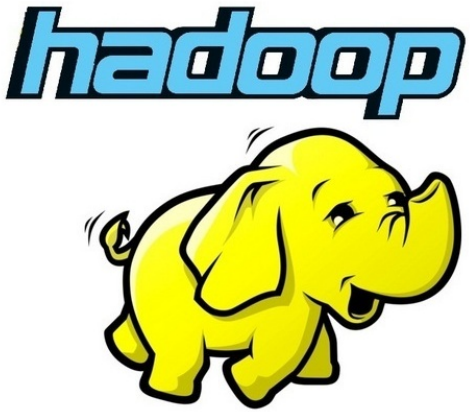
Courtesy of Google Trends: http://goo.gl/QHIQcN

HUMANGEO

Big data: De-mystified

What is big data?

What is Hadoop File System? (HDFS)



What is Hadoop MapReduce? (MR)

Source: Abe Usher

20th Century model of analysis

Tracy Morrow (aka “Ice T”)

How can you identify a legitimate hip-hop artist (versus someone who just gets up and rhymes)?

http://www.npr.org/2005/08/30/4824690/original-gangster-rapper-and-actor-ice-t

Source: Abe Usher

HUMANGEO

20th Century model of analysis

Tracy Morrow (aka "Ice T")

How can you identify a legitimate hip-hop artist (versus someone who just gets up and rhymes)?

"*Game knows game, baby.*"

Source: Abe Usher

HUMANGEO

20th Century model of analysis

Tracy Morrow (aka “Ice T”)

How can you identify a legitimate hip-hop artist (versus someone who just gets up and rhymes)?

“If you have expert knowledge, then you are capable of answering complex questions by interpreting domain specific information.” **[paraphrased]**

Source: Abe Usher

HUMANGEO

New model of analysis

Office Space

Peter Gibbons hatches a plot to write a computer virus that grab fractions of a penny from a corporate retirement account. http://goo.gl/rDg1U

Takeaway point: Little bits of value (information) provide deep insights in the aggregate

Source: Abe Usher

HUMANGEO

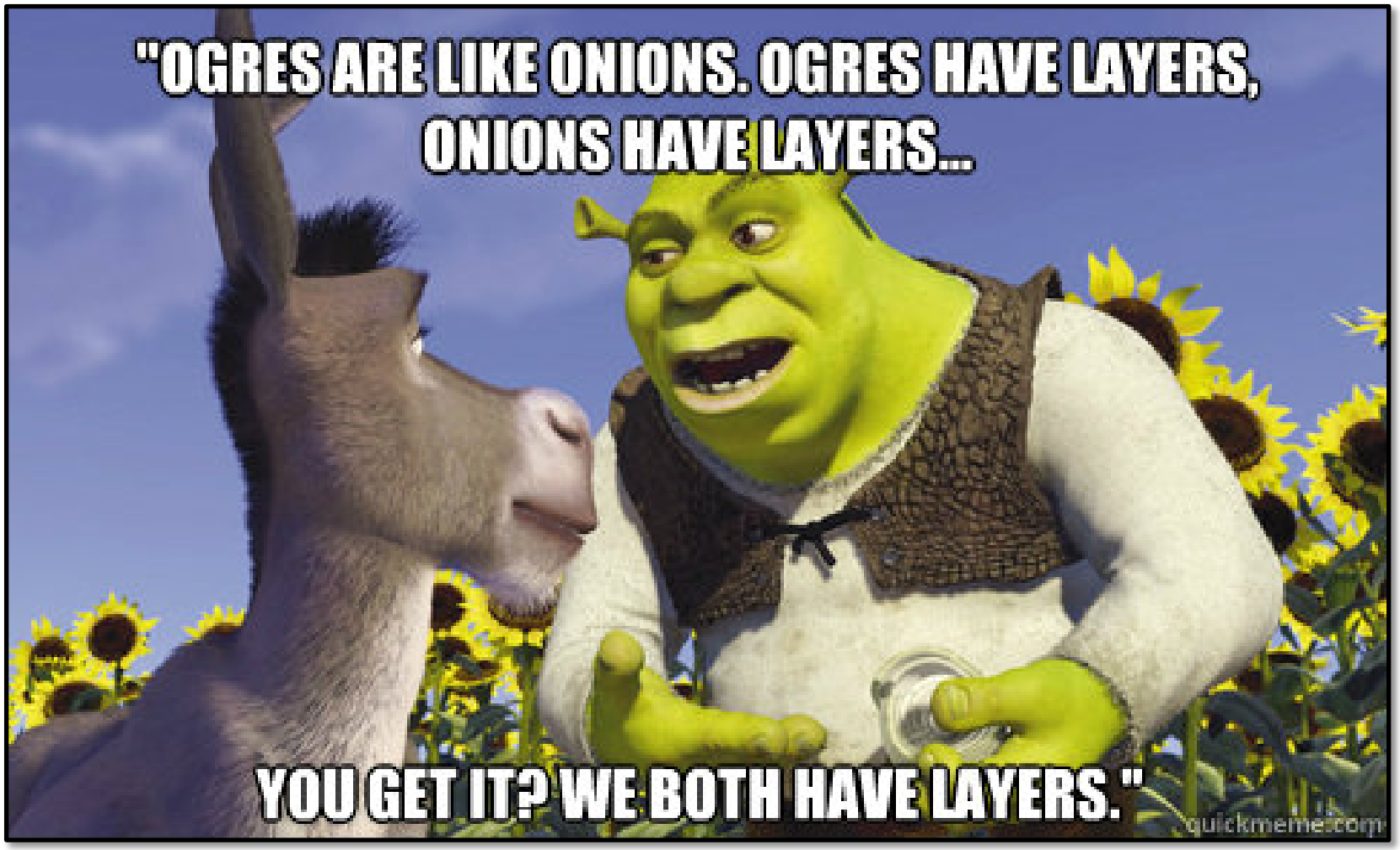
New model of analysis

Takeaway point: Hadoop simplifies the creation of massive counting machines

Source: Abe Usher

HUMANGEO

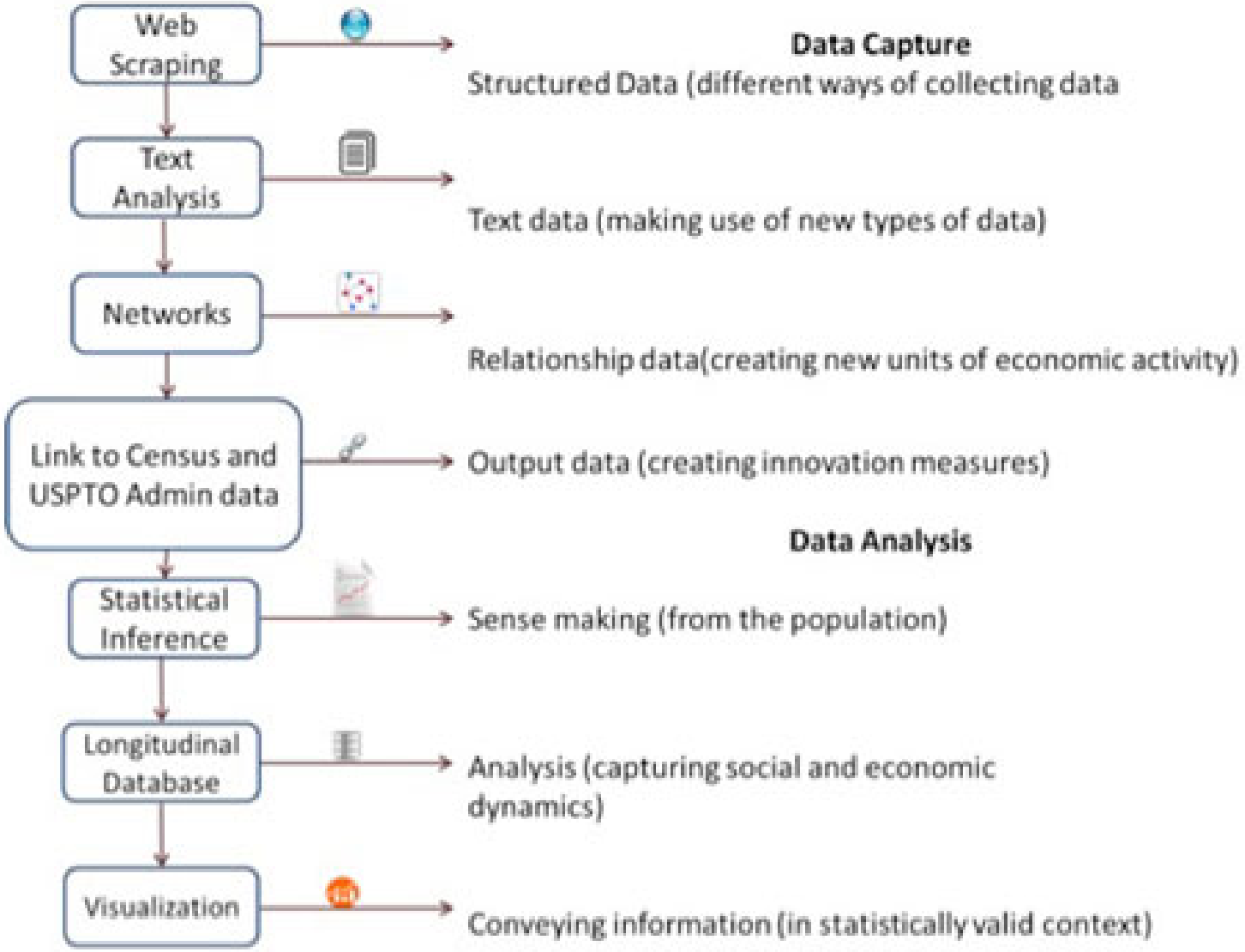
Big Data: Layers



Source: Abe Usher

HUMANGEO

Big Data: Layers

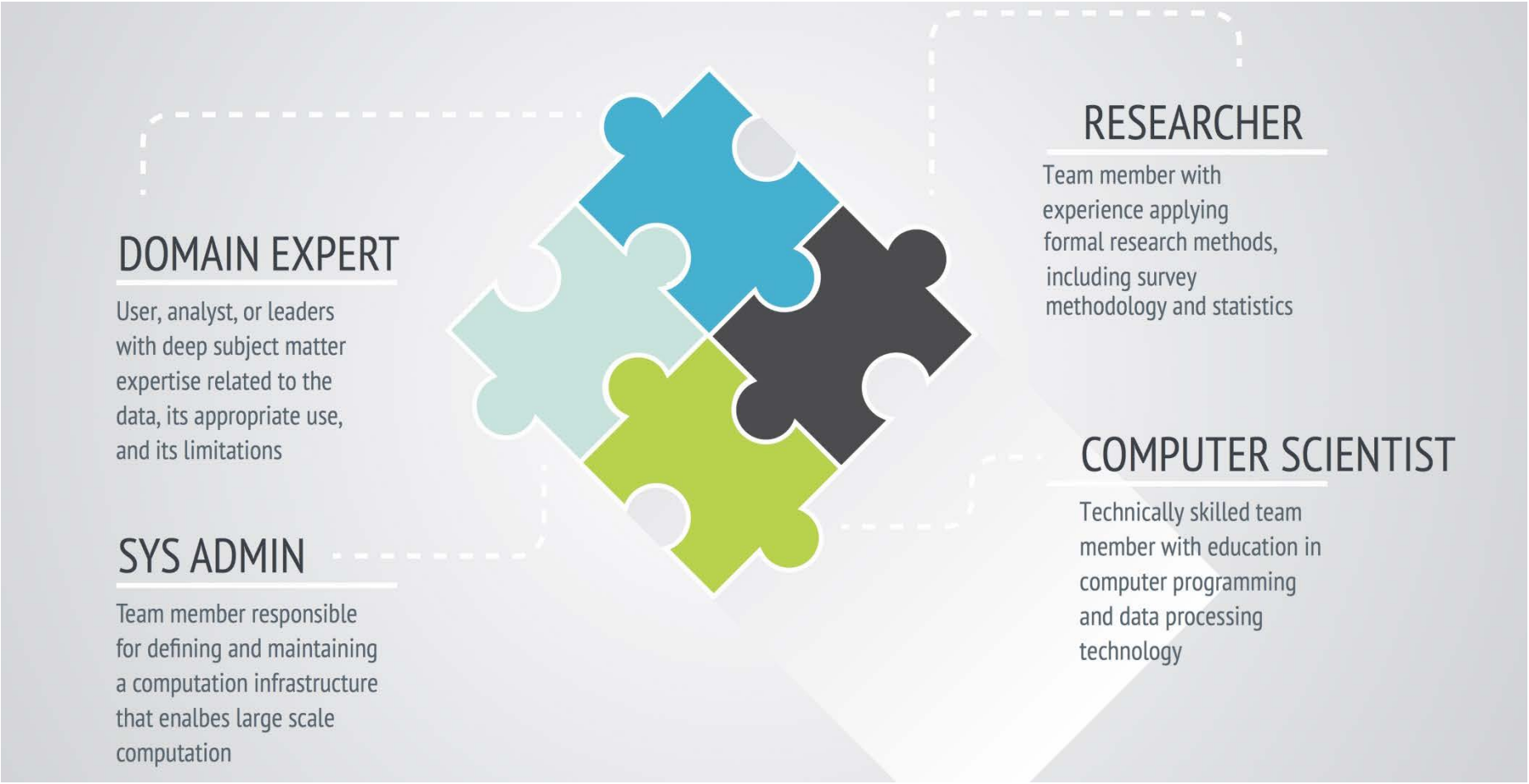


System

erest)

HUMANGEO

Four roles related to big data: each provide different skills



Source: Abe Usher

HUMANGEO

Skills by role

Computer scientist

- Data preparation
- MapReduce algorithms
- Python/R programming
- Hadoop ecosystem

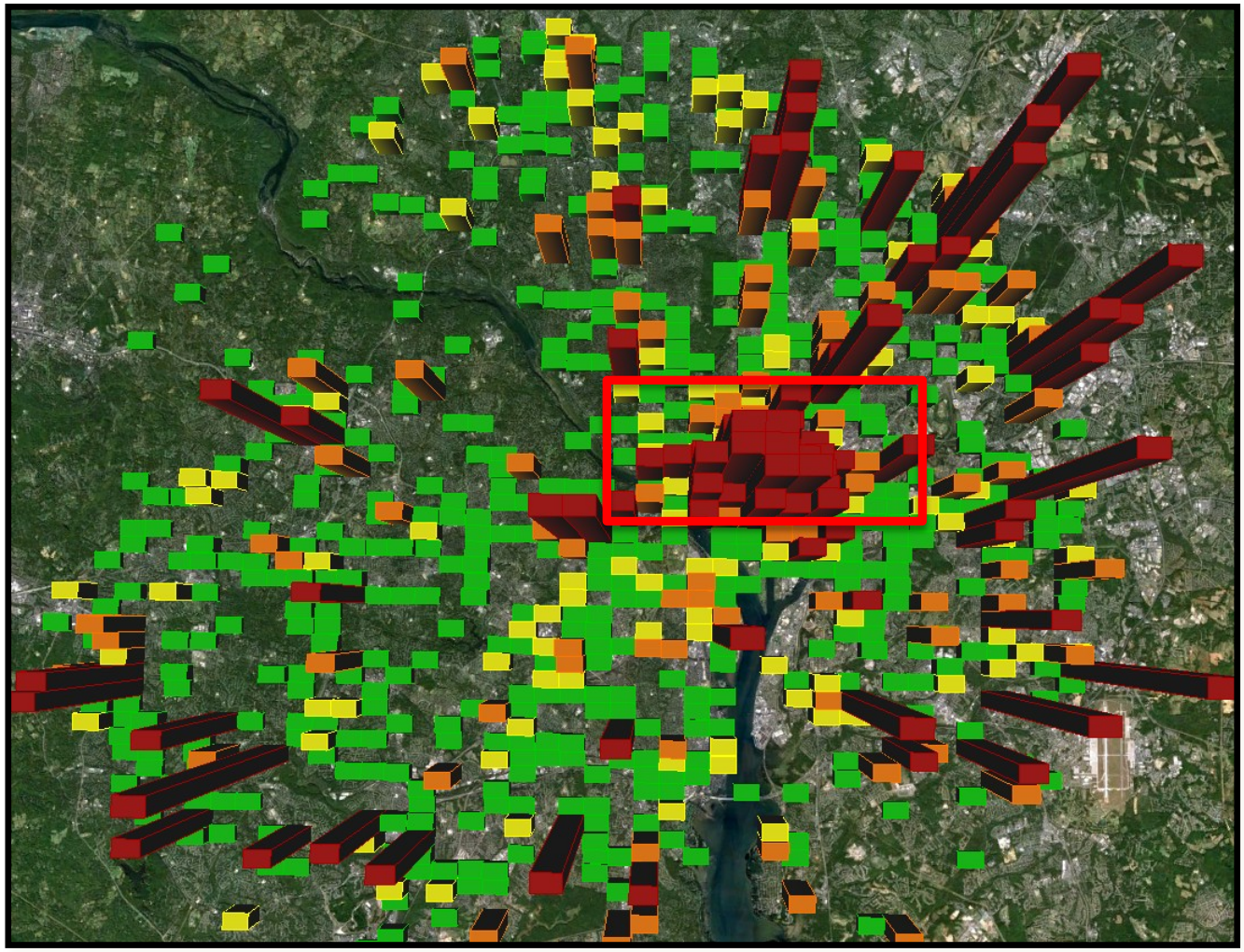
System Administrator

- Storage systems (MySQL, Hbase, Spark)
- Cloud computing:
  - Amazon Web Services (AWS)
  - Google Compute Engine
- Hadoop ecosystem

Source: Abe Usher

HUMANGEO

Big data enables new insights into human behavior



Geolocated social media activity in Washington DC during a 15 minute time period generated by MR. TweetMap

Source: Abe Usher

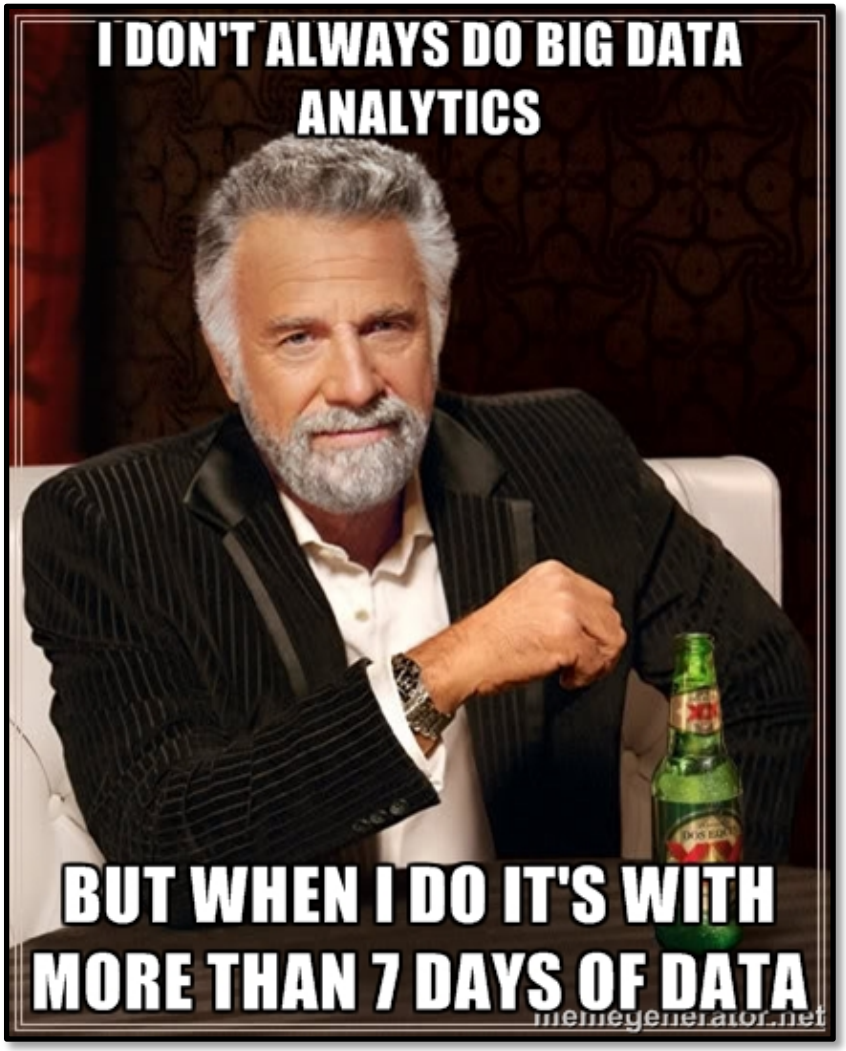
HUMANGEO

Contact information

Abe Usher

abe@thehumangeo.com

http://www.thehumangeo.com/



Easter Eggs

Google “do a barrel roll”

“Google gravity”

Google search in Klingon www.google.com/?hl=kn

Source: Abe Usher

RTI INTERNATIONAL | turning knowledge into practice

Big Data Veracity: Error Sources and Inferential Risks

Paul Biemer
RTI International
and
University of North Carolina

Source: Paul Biemer

Errors in Big Data: An Illustration

Suppose 1 in 1,000,000 people are terrorists

The Big Data Terrorist Detector is 99.9 accurate

The detector says your friend, Jack is a terrorist.

What are the odds that Jack is really a terrorist?

Source: Paul Biemer

Errors in Big Data: An Illustration

Suppose 1 in 1,000,000 people are terrorists

The Big Data Terrorist Detector is 99.9 accurate

The detector says your friend, Jack is a terrorist.

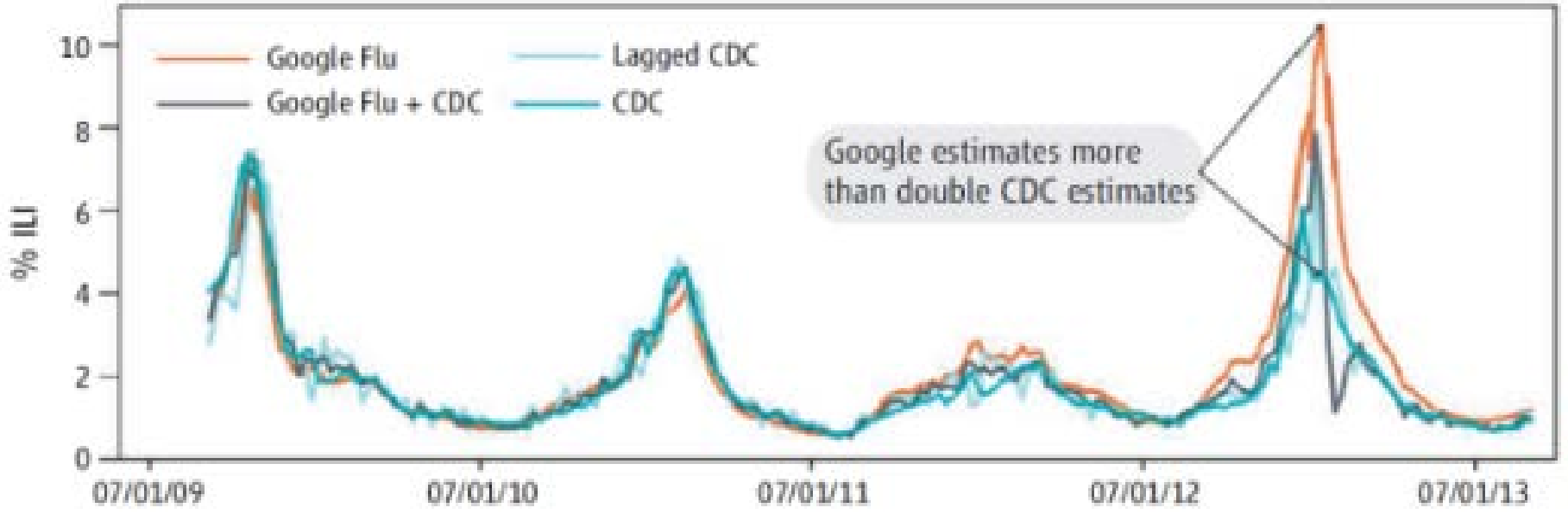
What are the odds that Jack is really a terrorist?

Answer: 1 in 1000 i.e.,
99.9% of the terrorist detections will be false!

Source: Paul Biemer

Some questions regarding Big Data veracity

- What constitutes a Big Data error?



Systematic error in Google Flu Trends data

Source: Paul Biemer

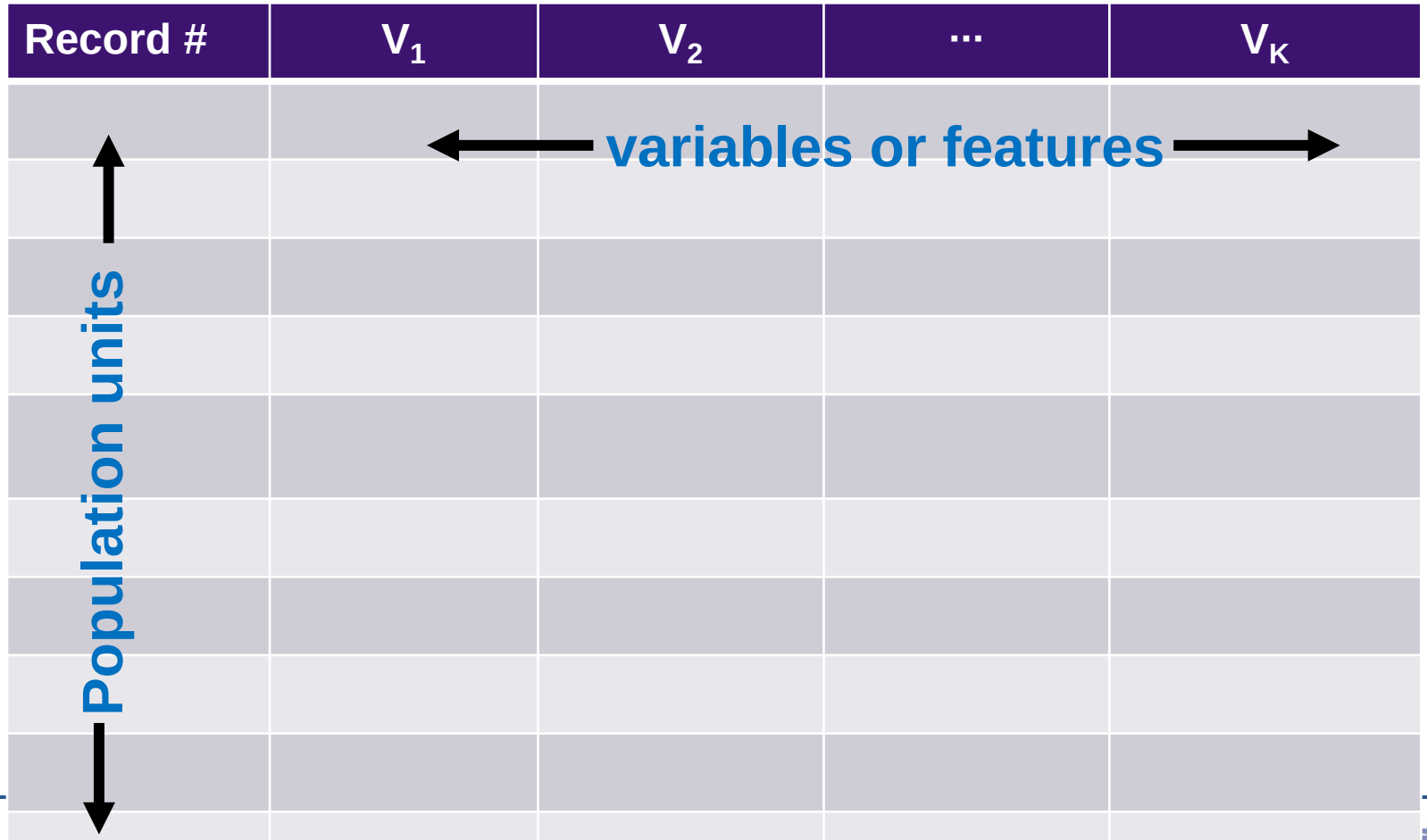
Some questions regarding Big Data veracity

- What constitutes a Big Data error?
- What are the sources and causes of the errors?
- Do the error distributions vary by source?
- Are the errors systematic or variable or both?
- How do the errors affect data analysis such as
  - Classifications
  - Correlations
  - Regressions
- How can analysts mitigate these effects?

Source: Paul Biemer

Total Error Framework for Traditional Data Sets

Typical File Structure

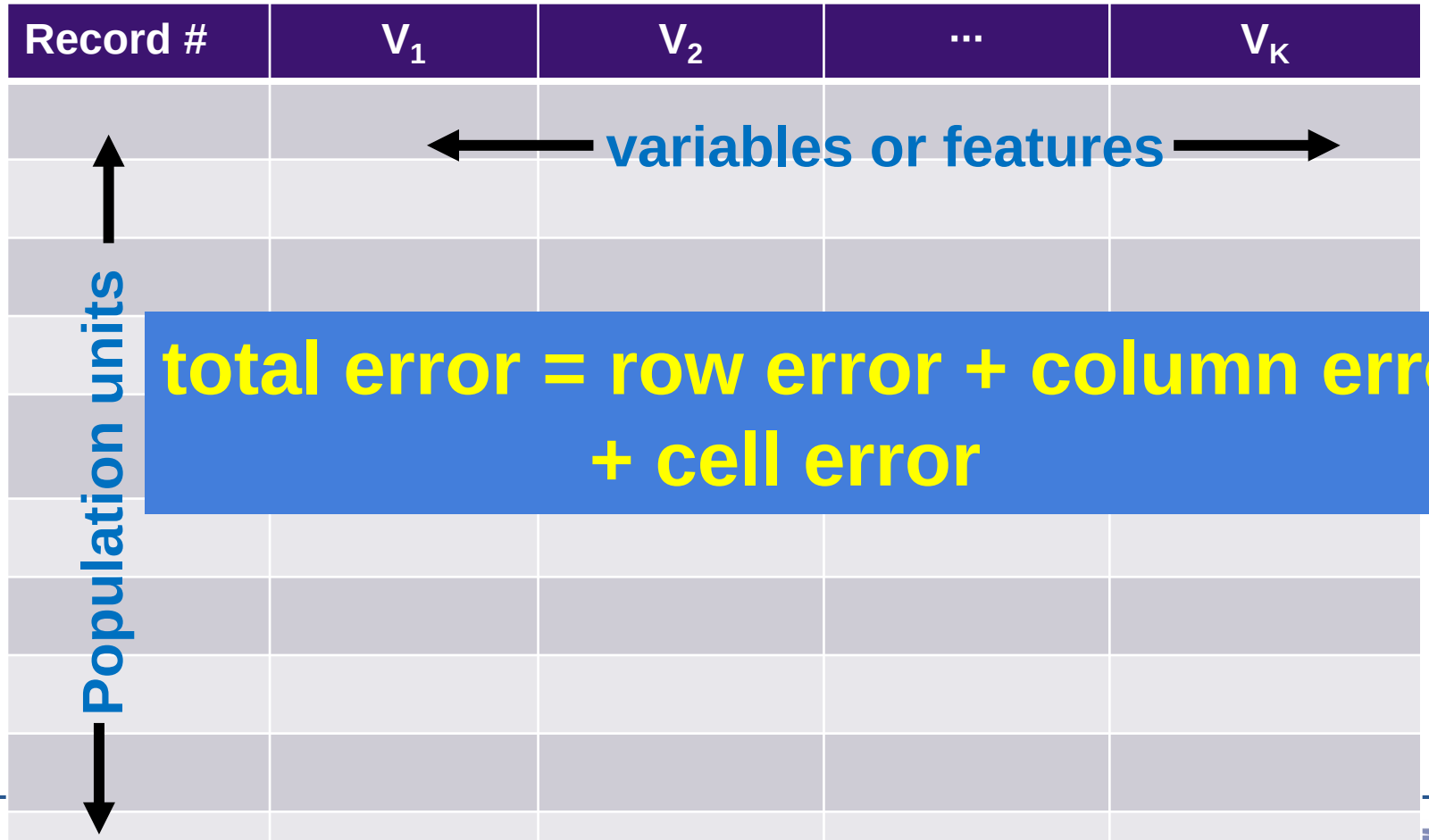


Source: Paul Biemer

RTI INTERNATIONAL

Total Error Framework for Traditional Data Sets

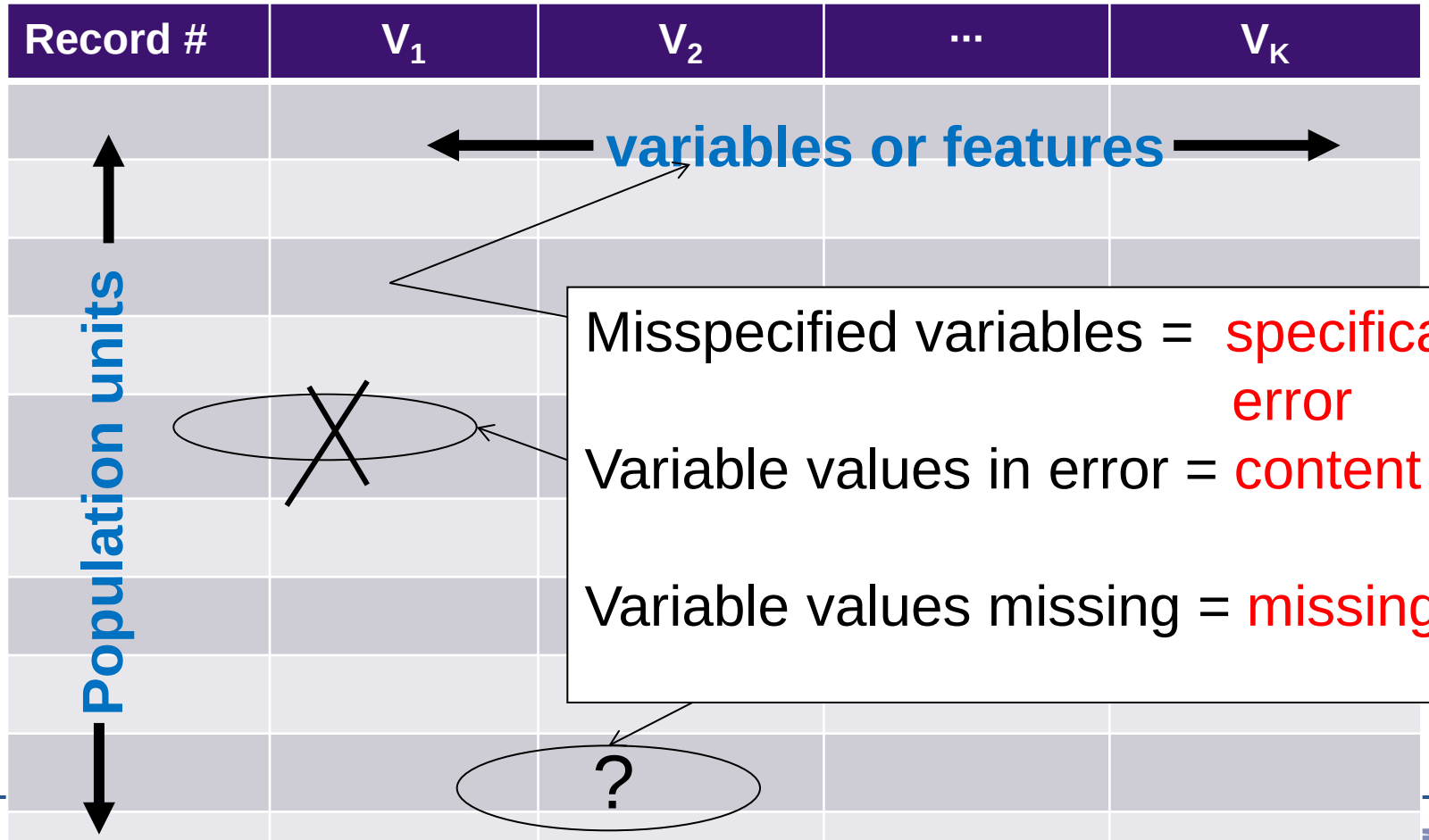
Typical File Structure



Source: Paul Biemer

Possible Column and Cell Errors

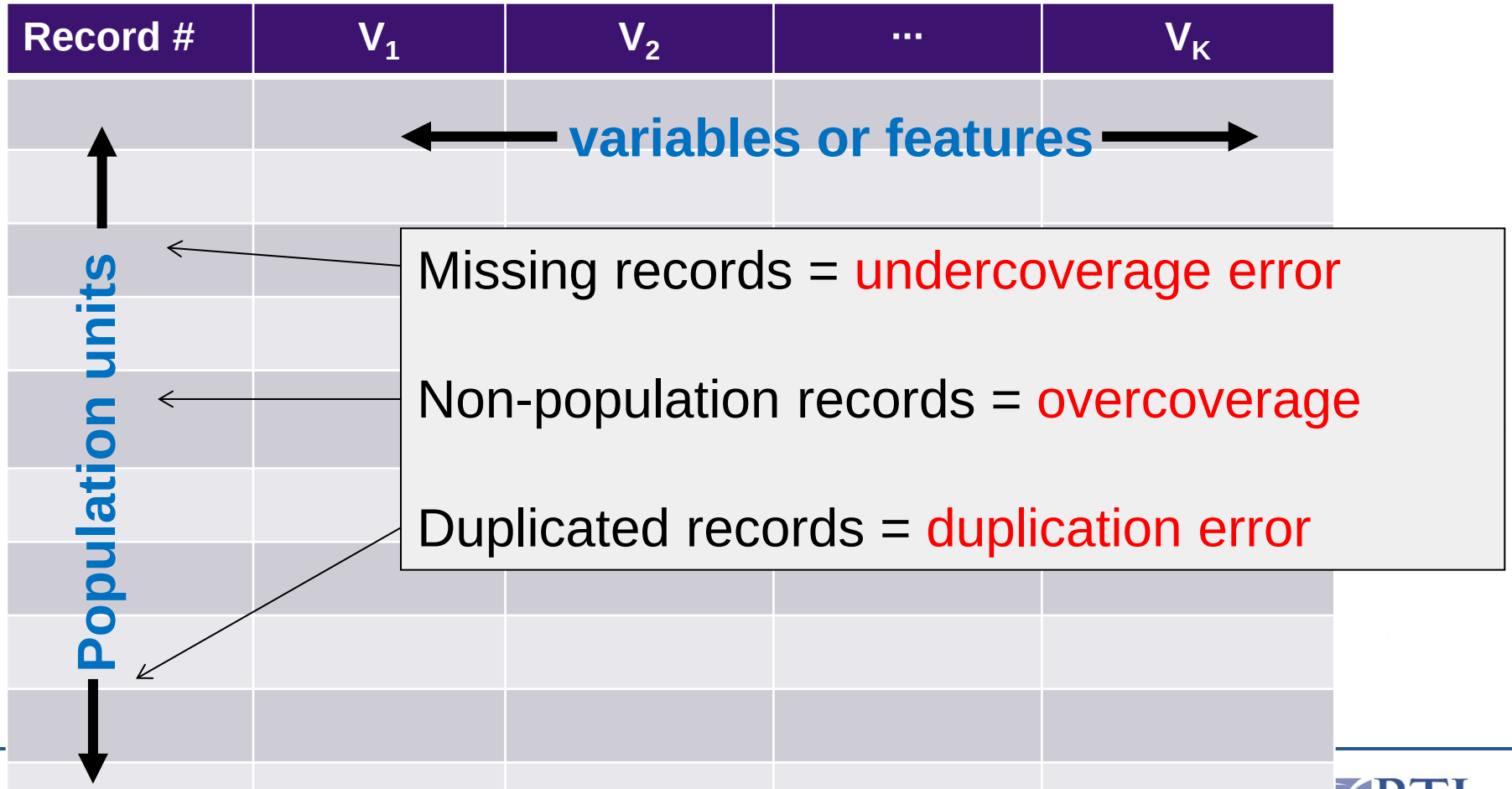
Typical File Structure



Source: Paul Biemer

Possible Row Errors

Typical File Structure



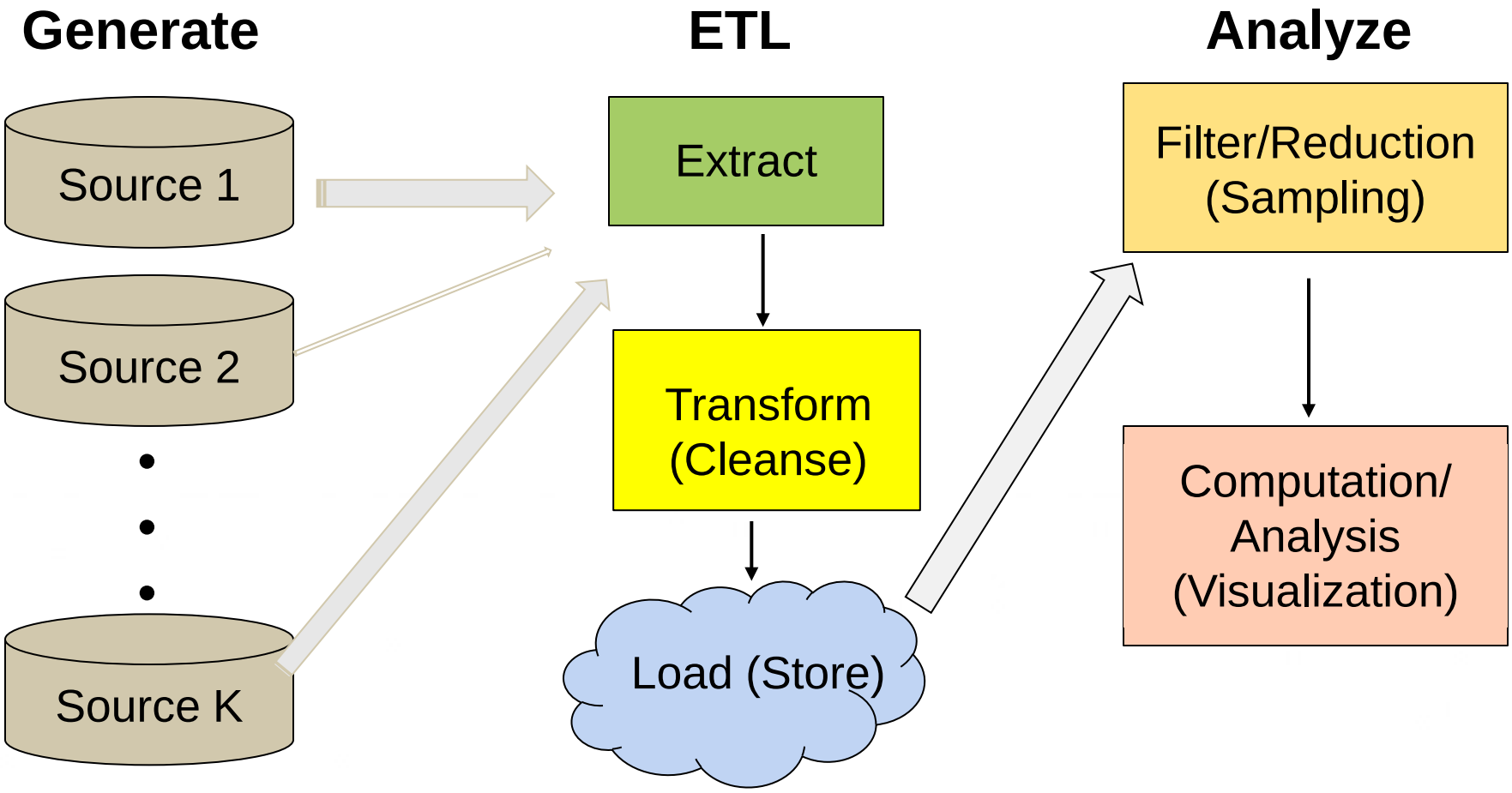
Source: Paul Biemer

Shortcomings of the Traditional Framework for Big Data

- Big Data files are often not rectangular
  - hierarchically structure or unstructured
- Data may be distributed across many data bases
  - Sometimes federated, but often not
- Data sources may be quite heterogeneous
  - Includes texts, sensors, transactions, and images
- Errors generated by Map/Reduce process may not lend themselves to column-row representations.

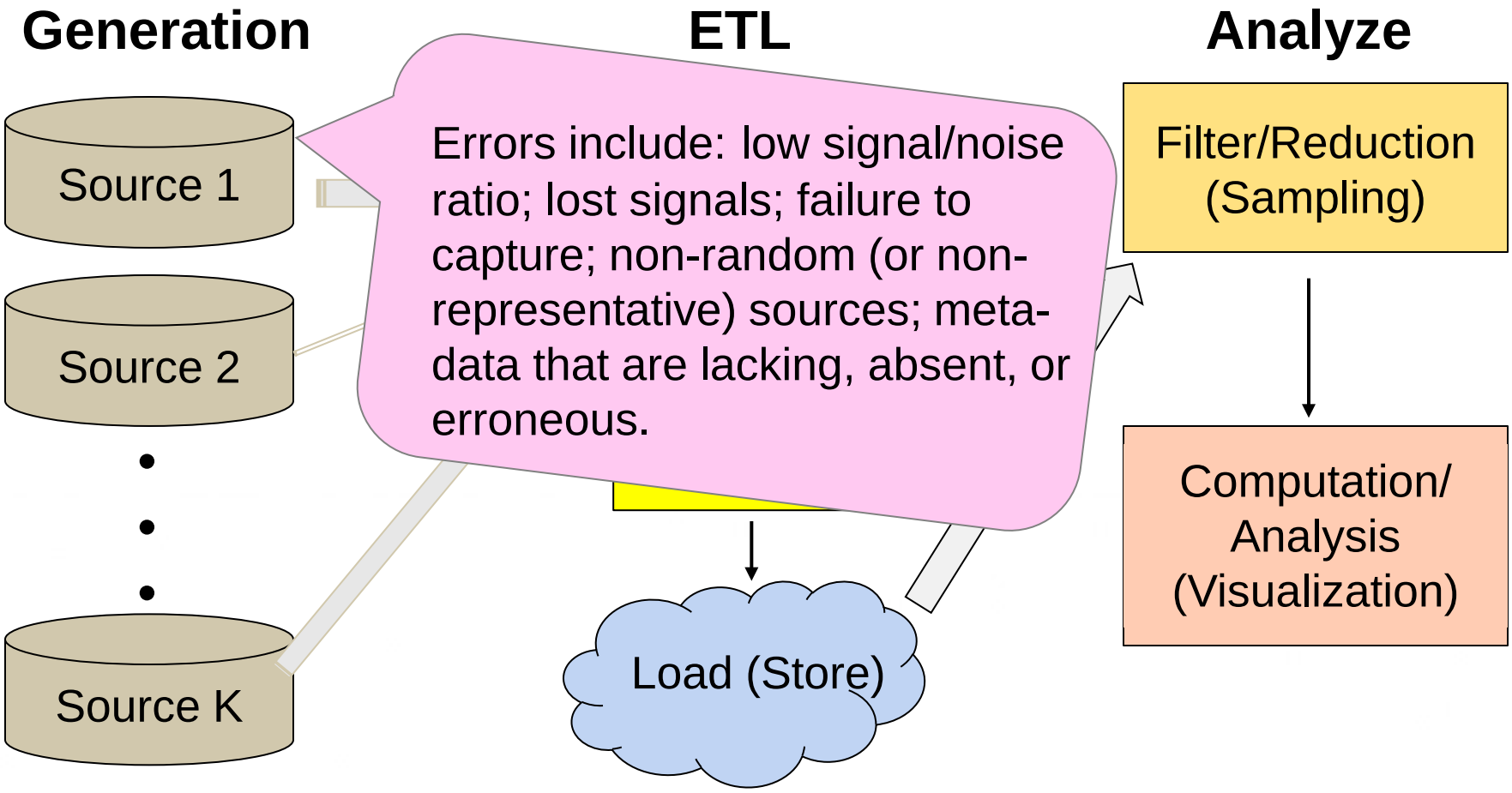
Source: Paul Biemer

Big Data Process Map



Source: Paul Biemer

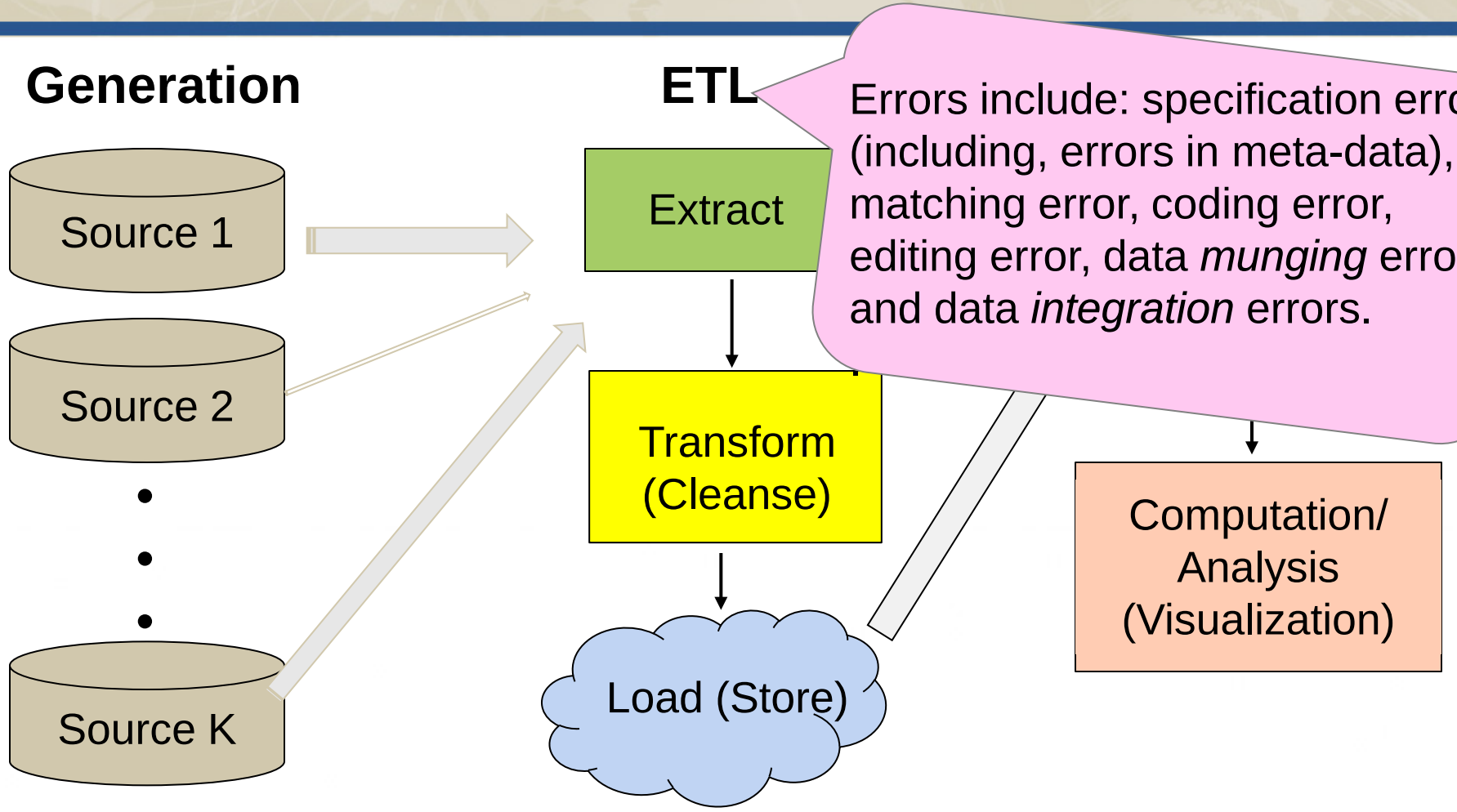
Big Data Process Map



Source: Paul Biemer

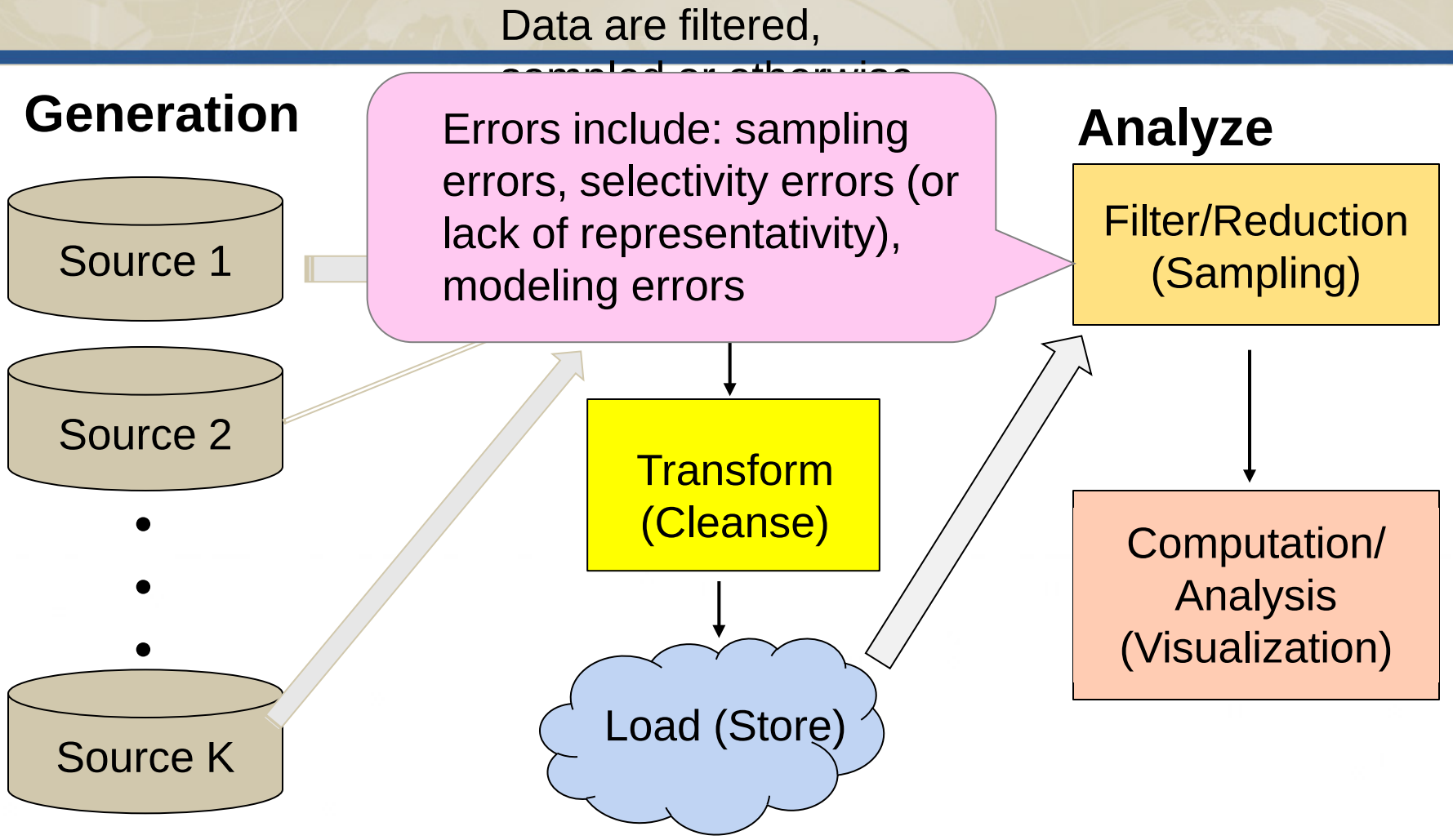
RTI INTERNATIONAL

Big Data Process Map



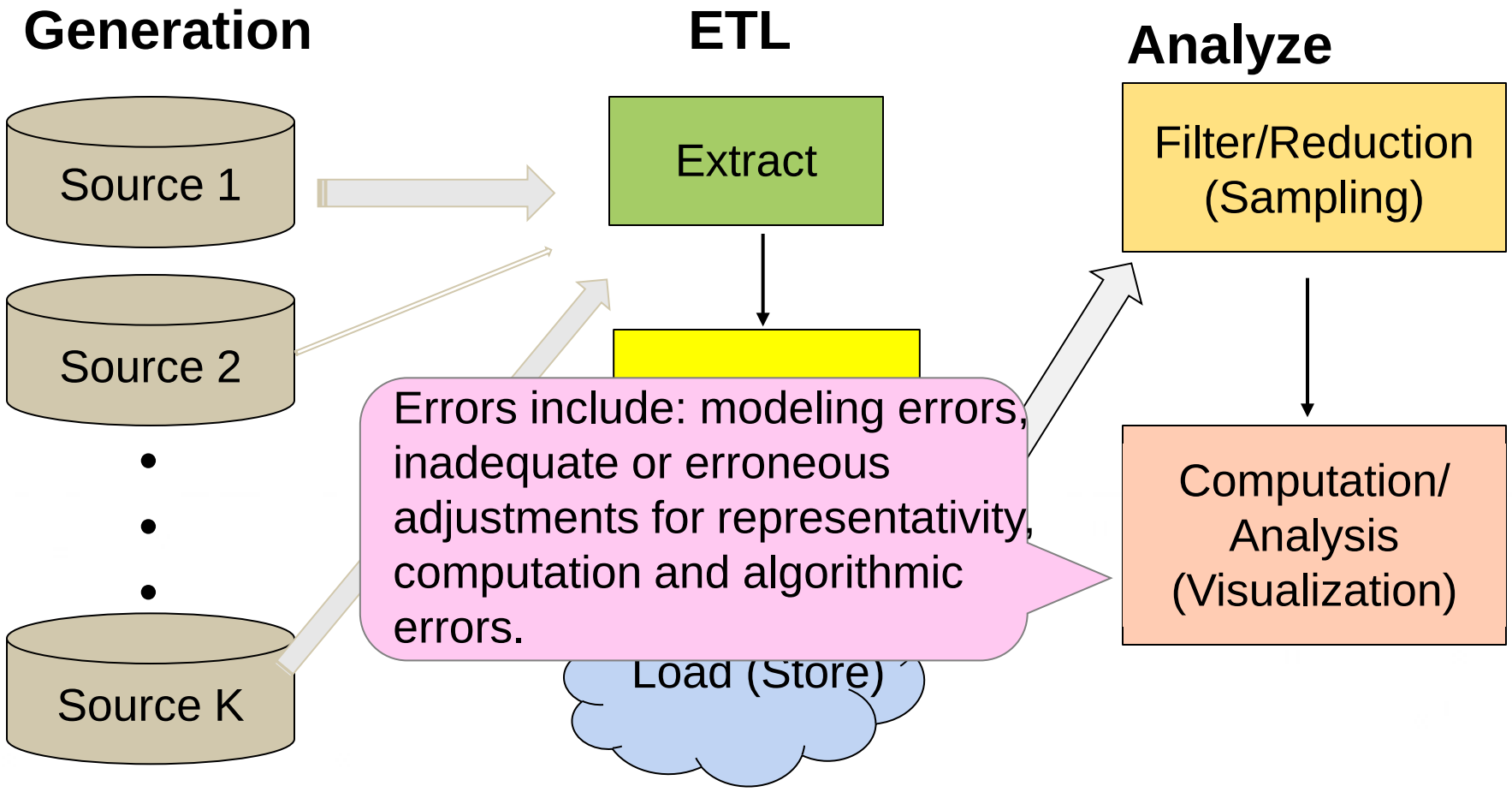
Source: Paul Biemer

Big Data Process Map



Source: Paul Biemer

Big Data Process Map

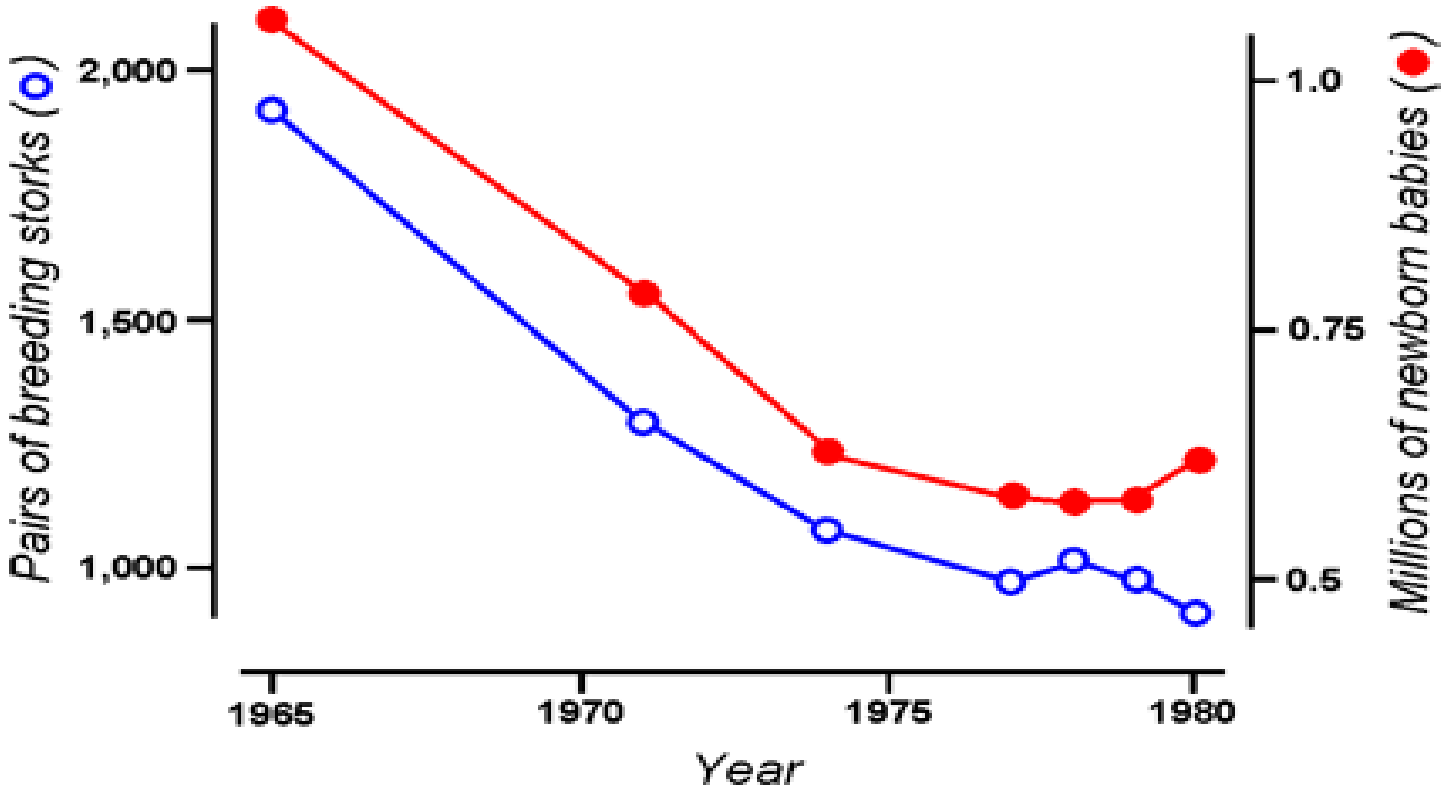


Source: Paul Biemer

RTI INTERNATIONAL

Implications for Data Analysis

Stork Die-off Linked to Human Birth Decline



Source: Paul Biemer

Implications for Data Analysis

- Study of rare groups is problematic
- Biased correlational analysis
- Biased regression analysis
- Coincidental correlations
- Noise accumulation – inability to identify correlates
- Incidental endogeneity – Cov(error, covariates)

Source: Paul Biemer

Implications for Data Analysis

- Study of rare groups is problematic
- Biased correlational analysis
- Biased regression analysis
- Coincidental correlations
- Noise accumulation – inability to identify correlates
- Incidental endogeneity – Cov(error, covariates)
  - These latter three issues are a concern even if the data could be regarded as error-free.
  - Data errors can considerably exacerbate these problems.

Current research is aimed at investigating these errors.

Source: Paul Biemer

Recommendations

1. *Surveys and Big Data are complementary data sources not competing data sources. There are differences between the approaches, but this should be seen as an advantage rather than a disadvantage.*
2. *AAPOR should develop standards for the use of Big Data in survey research when more knowledge has been accumulated.*

3. AAPOR should start working with the private sector and other professional organizations to educate its members on Big Data

4. AAPOR should inform the public of the risks and benefits of Big Data.

5. AAPOR should help remove the barrier associated with different uses of terminology.

6. AAPOR should take a leading role in working with federal agencies in developing a necessary infrastructure for the use of Big Data in survey research.

CRA Data Program

Patricia Whitridge, Canada Revenue Agency
SOI Panel Meeting - June 5,2015

Why the need for a CRA Data Program?

- Better deal with needs for data that cross program and organizational boundaries, and external data
- Currently, no single entity responsible for the data – ITB managed in past
- Provide leadership in new types of data usage (e.g. analytics) and new data directions (e.g. Open Data)
- Provide a coordinated whole-of-Agency approach to acquiring, using, sharing, managing, and publishing data

Program Scope

- Builds on existing functions:
  - Data provisioning services (per data sharing agreements, and ad hoc requests)
  - Business Intelligence Centre of Expertise
  - Maintain data catalogue, supported by Managed Metadata Environment
  - Data Stewardship

- New functions
  - Identify new data sources and opportunities
  - Data governance
  - Data policies and standards (business-oriented)
  - Talent management
  - Oversee – and report on – state of Agency's data

Proposed Program Model

- Centralized leadership and coordination, with cross-Agency involvement
- Existing functional and technical responsibilities for managing data are maintained
- Data Program works in partnership with branches/regions and ITB
- Accountability for Data Program is assigned to a new role: Chief Data Officer

LEI and Big Data

Arthur Kennickell
Federal Reserve Board
SOI Panel Meeting
June 5, 2015

Opinions expressed are those of the presenter and do not necessarily reflect the views of the Federal Reserve or its staff.

LEI: Review and Update

- LEI: identification system for entities
  - "Entities" broadly construed
  - ISO 17442: 2012 standard serves as basis
- Fixed identification number
  - Unique code: 20 characters, including 2 check digits
    - Example: U.S. Sugar Corp.: **5493***00*VV3SF28T3NK5**85**
  - Exclusive assignment
  - Persistent
  - No persistent embedded intelligence
- Reference data
  - Variable data
  - Data to support precise identification of individual entities

LEI Scope

- ISO 17442: “The term legal entities includes, **but is not limited to** unique parties that are legally or financially responsible for the performance of financial transactions or have the legal right in their jurisdiction to enter independently into contracts…”
  - Excludes natural persons acting as natural persons
  - Otherwise very broad, in principle—even government entities
  - Many edge cases: quasi-entities, contract-based structures, large individual actors in markets
  - In practice, proceeding cautiously to avoid confusion
- On-going work with ISO to define a standard for legal form
  - Could be used to define eligibility for LEI

Global LEI System (GLEIS)

- Regulatory Oversight Committee (ROC)
  - Committee on Evaluation and Standards (CES)
- Global LEI Foundation (GLEIF)
- Local Operating Units (LOUs)
- Registrants
- Local regulators
- Users

Funding Model

- Use is free
- Nonprofit, cost-recovery principles
- Charge for registration and annual maintenance
  - Initial charge currently about $200
  - Maintenance fee about $100
  - Scale economies expected to drive down cost
- A fixed fee charged to LOUs for each registered LEI supports the GLEIF

Data Quality

- Highest priority: **Information must be reliable and timely**
- Validation
  - Entity exists, person applying authorized, entity within scope
  - Reference data confirmed
  - Responsibility for accuracy rests primarily with the registrant
- Public challenge facility

Data on Organizational Relationships

- Relationship data needed for aggregating exposures or for tracing flows of money or information
  - Essential both for regulators and private risk managers to aggregate disparate sources of data to understand risk
  - Will also support broader transparency
    - E.g.. Anti money laundering
  - First relationships considered: “Direct parent”/”Ultimate parent”
    - Need for extensible structure to accommodate other relationships in the future
  - Other elements of relationship need to be considered
    - Numerous technical complications
    - Privacy issues and jurisdictions with opaque ownership structures

First Phase of Organizational Relationship Data

- Based on accounting consolidation definitions
- Open questions
  - Collect from “parents” or “children”—or both
  - Roles of the LOUs and the GLEIF in collecting/consolidating data
  - Timeliness
  - How encourage compliance in absence of universal regulatory mandate
  - How to address “holes” in organizational structures in the short run
    - Both “missing LEIs” for one side of relationship and entire missing branches of organizations
  - Appropriate sources of data for validation
- Iterative proof of concept planned
- On-going consultation with regulators and private sector
- First phase of implementation expected around end of 2015

Standardization

- LEI is a classic example of the role of standardization in making big data useable
  - ISO TC 68 deals with all standards for the financial industry
- Other standards bodies
- Broad role for standardization
- Importance of metadata
  - “local standardization”
  - Data point modeling

“Big data hubris”

- Big Data: Are we making a big mistake? (Tim Harford in the *Financial Times*)
  - Often clash with accepted statistical procedures
  - Informative sampling, **nonstationarities**, ambiguous definitions/frame of reference, fuzzy provenance, other little-data problems writ large, etc
  - Are we looking for scientific results or are we surfing on short-horizon buzz?

- Almost certainly not a mistake in general, but:

There is no substitute for thinking about what you are doing

Are we all Bayesians now?

- Modeling, matching, simulating
- “Priors” over unknown population definitions or data content?

- SOI data may provide a key universe anchor for other parts of the “data collage”

Big Data Skills Mix

- **Domain expert**
  - Traditional strength of that SOI embodies as an organization
  - Data curation and matching
  - Also able to do other things: role in analytics
- Researcher (methodology, statistics, mathematics)
- Computer scientist
- System administrator

- Significant interdependencies

Thanks!

Panel Discussion

Thinking Big About SOI Data

Discussion Questions

- Given current resources, what changes in current products or production methods should SOI consider in order to free resources for new work?
- What changes to public data releases should be developed using linked flow-through data?

Lunch:

1 hour 15 minutes

Next:

Are Piketty and Zucman Getting it Right?

2015 Consultants Panel Agenda (afternoon sessions)

Are Piketty and Zucman Getting it Right? Evaluating Distributional Statistics Based on Aggregate Data		
More Than They Realize: The Income of the Wealthy and the Piketty Thesis	Jenny Bourne	1:20 pm
Measuring Income at the Top	John Sabelhaus	
Mortality Differentials - How Much Longevity Can Money Really Buy?	Brian Raub	
Discussant	Len Burman	
Discussion	Panel	
A Productive Partnership, Joint Work with Stanford	David Grusky	2:30 pm
Discussion	Panel	
An Overview of the SOI Consultants Panel	George Plesko	3:10 pm
Discussion	Panel	
Adjorn		

ARE PIKETTY AND ZUCMAN GETTING IT RIGHT?

More Than They Realize: The Income of the Wealthy and the Piketty Thesis

Jenny Bourne, Carleton College, Economics Department

Eugene Steuerle and Ellen Steele, Urban Institute

Brian Raub and Joseph Newcomb, Statistics of Income, IRS

Measuring Inequality

Realized Capital Income ➔ Wealth **?**

The Wealthy Differ from You and Me

- Higher economic returns (portfolio holdings, selection bias)
- Lower realized returns (ability to re-categorize and time income)
- Effective tax rate on economic income from capital <10%

Data

Decedents from 2007 who filed estate tax return Form 706

(total N=36,889; stratified sample N=12,296)

Federal tax returns for 2002-2007 Form 1040

Total gross estate $229 billion

Mean gross estate $6.2 million, mean net estate $6 million

Median gross estate $3.2 million, median net estate $3.15 million

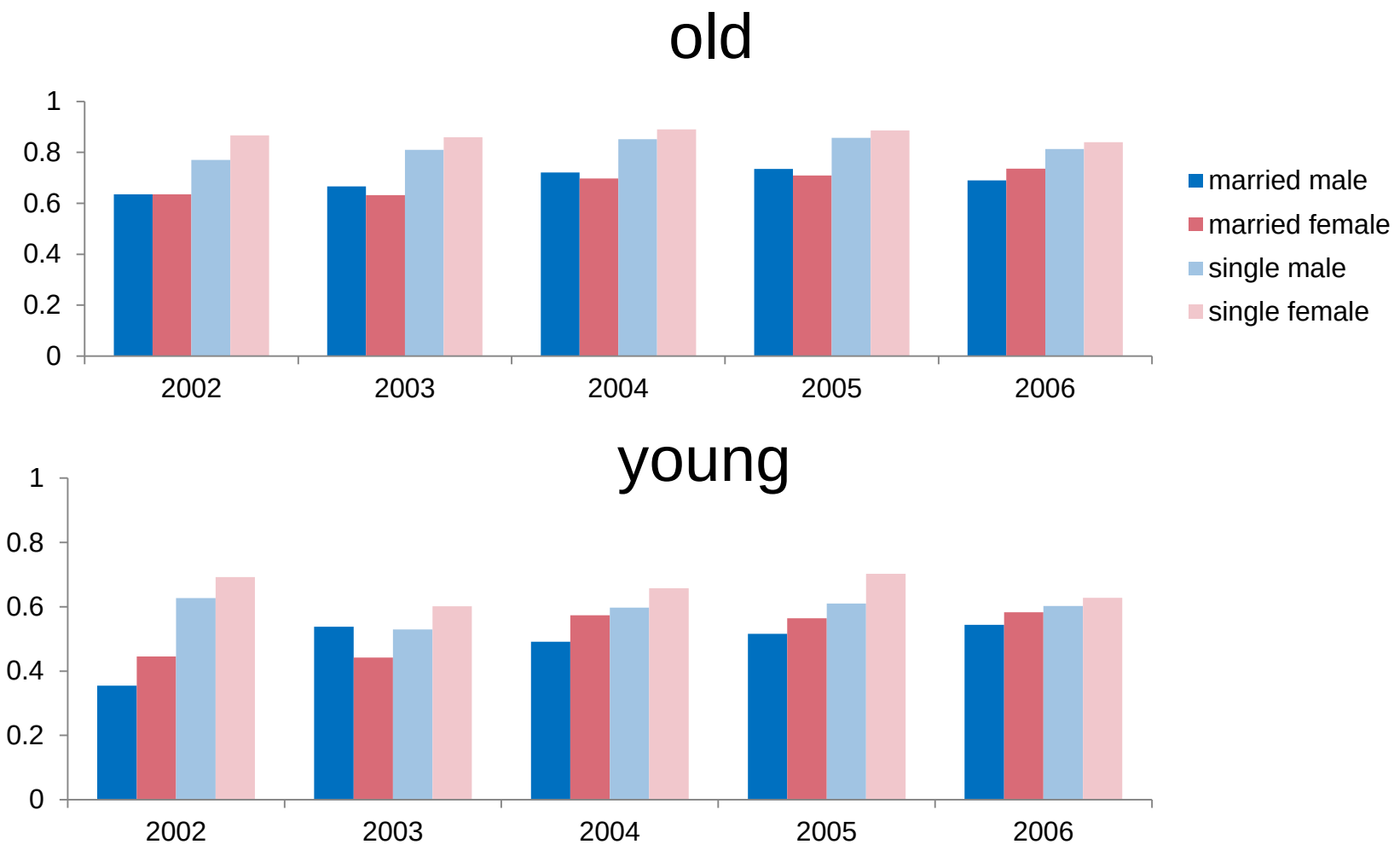
Descriptive Information

Demographic trait	
Age at death 70 years or older	80.3
Male	57.0
Married at death	49.3
Marital status 2002--2007	
Always married	45.6
Always not married	38.8
Mixed	5.6
Net estate category ($million)	
0--2	3.5
2--5	70.9
5--10	17.1
10--50	7.7
50--100	0.5
Over 100	0.3

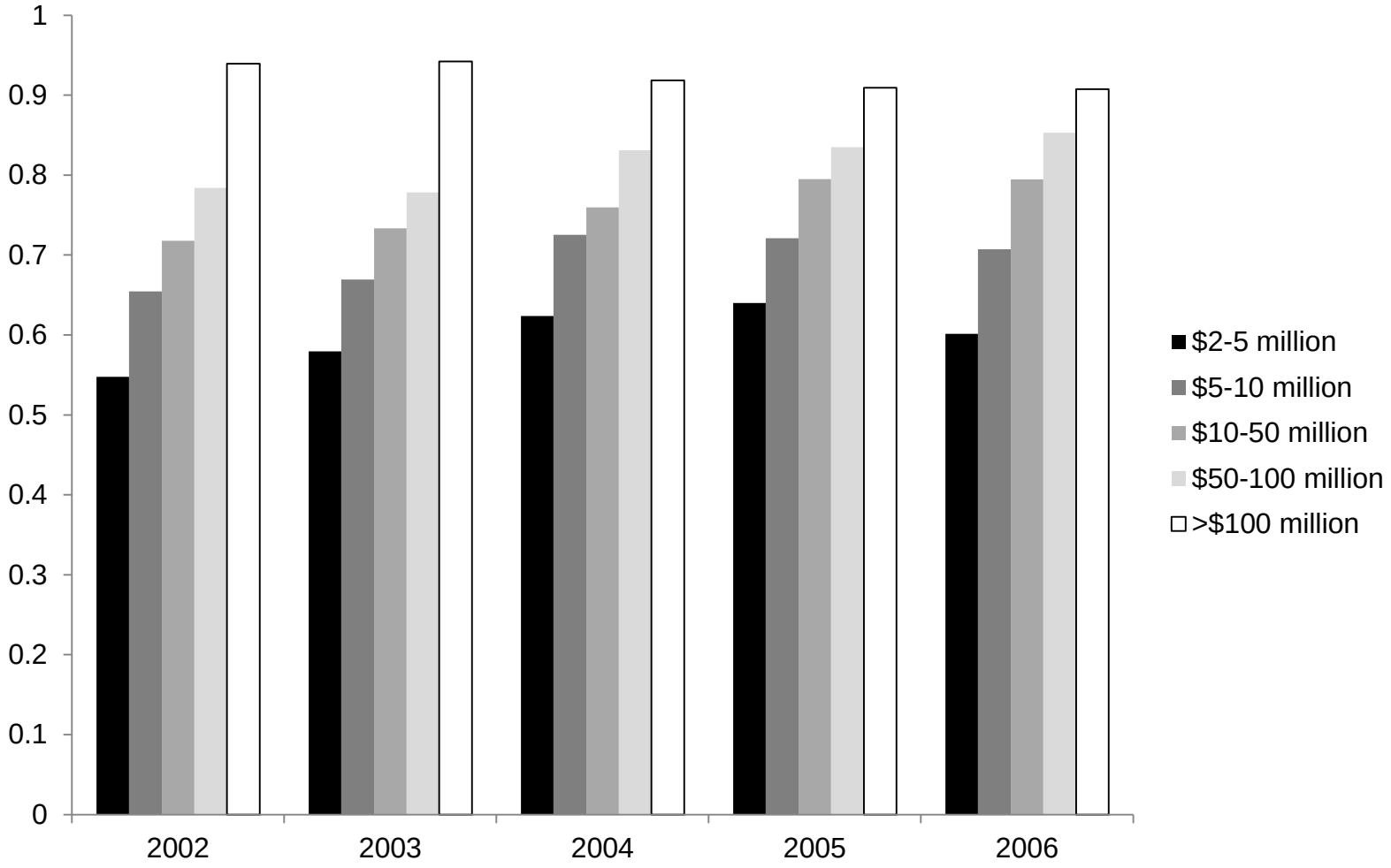
Three Measures of Capital Income

CAPY1	CAPY2	TAXY
Taxable interest	Taxable interest	Taxable interest
+Tax-exempt interest	+Tax-exempt interest	
+Capital gains	+Capital gains	+Taxable capital gains
+Dividends	+Dividends	+Dividends
+Gains from sale of business property	+Gains from sale of business property	+Gains from sale of business property
+½ Schedule C	+½ Schedule C	+½ Schedule C
+¾ Schedule E	+¾ Schedule E	+¾ Schedule E
+½ Schedule F	+½ Schedule F	+½ Schedule F
	+½ IRA distribution	
	+½ Pensions&annuities	
-Interest deduction	-Interest deduction	-Interest deduction

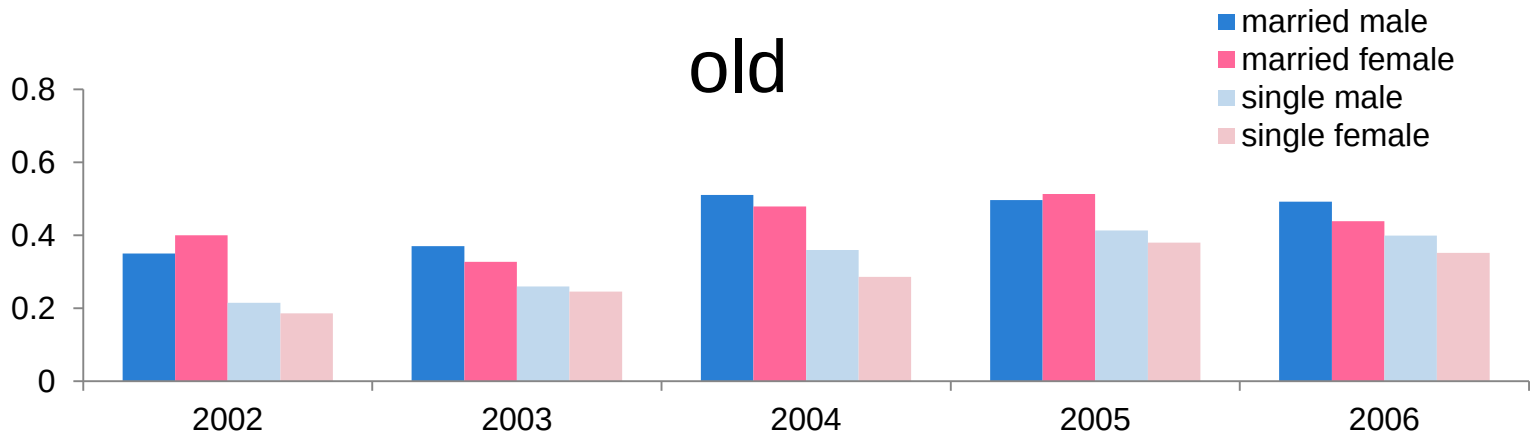
CAPY1 as a Proportion of Total Income, by Status (2002-2006)

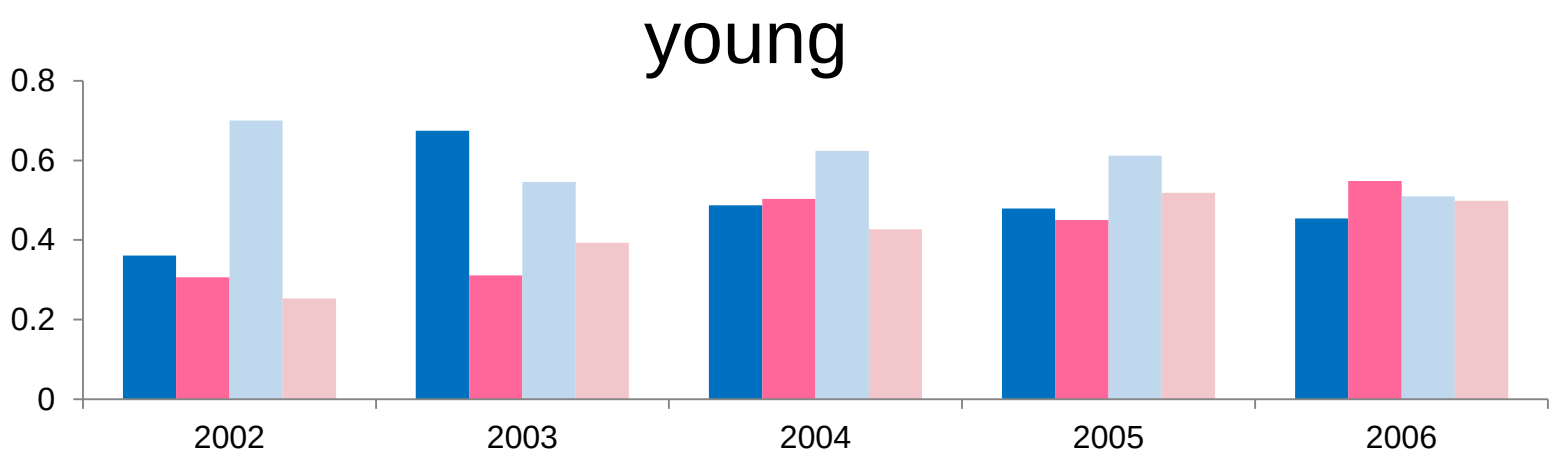


CAPY1 as a Proportion of Total Income, by Wealth Category (2002-2006)

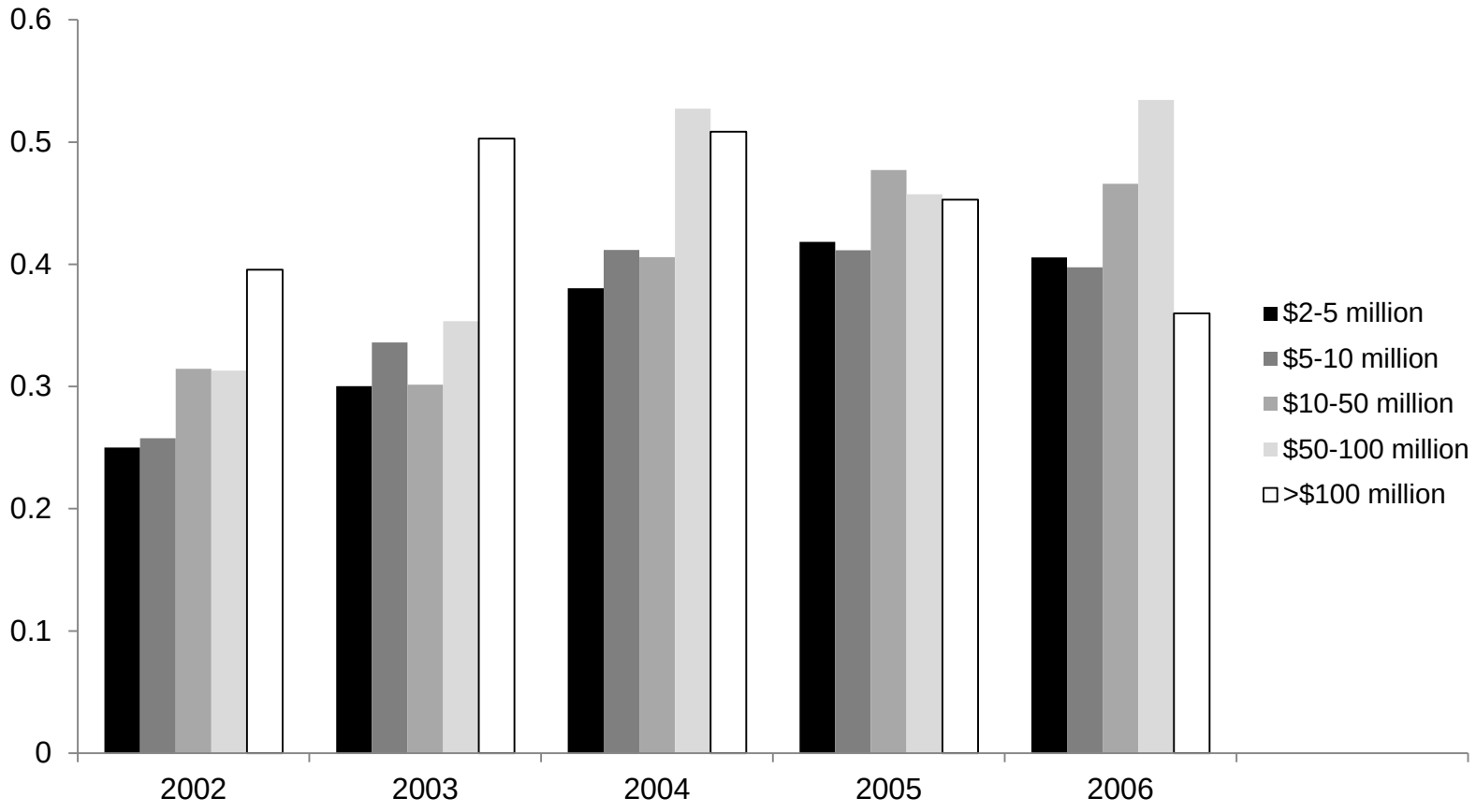


Capital Gains as a Proportion of CAPY1, by Status (2002-2006)

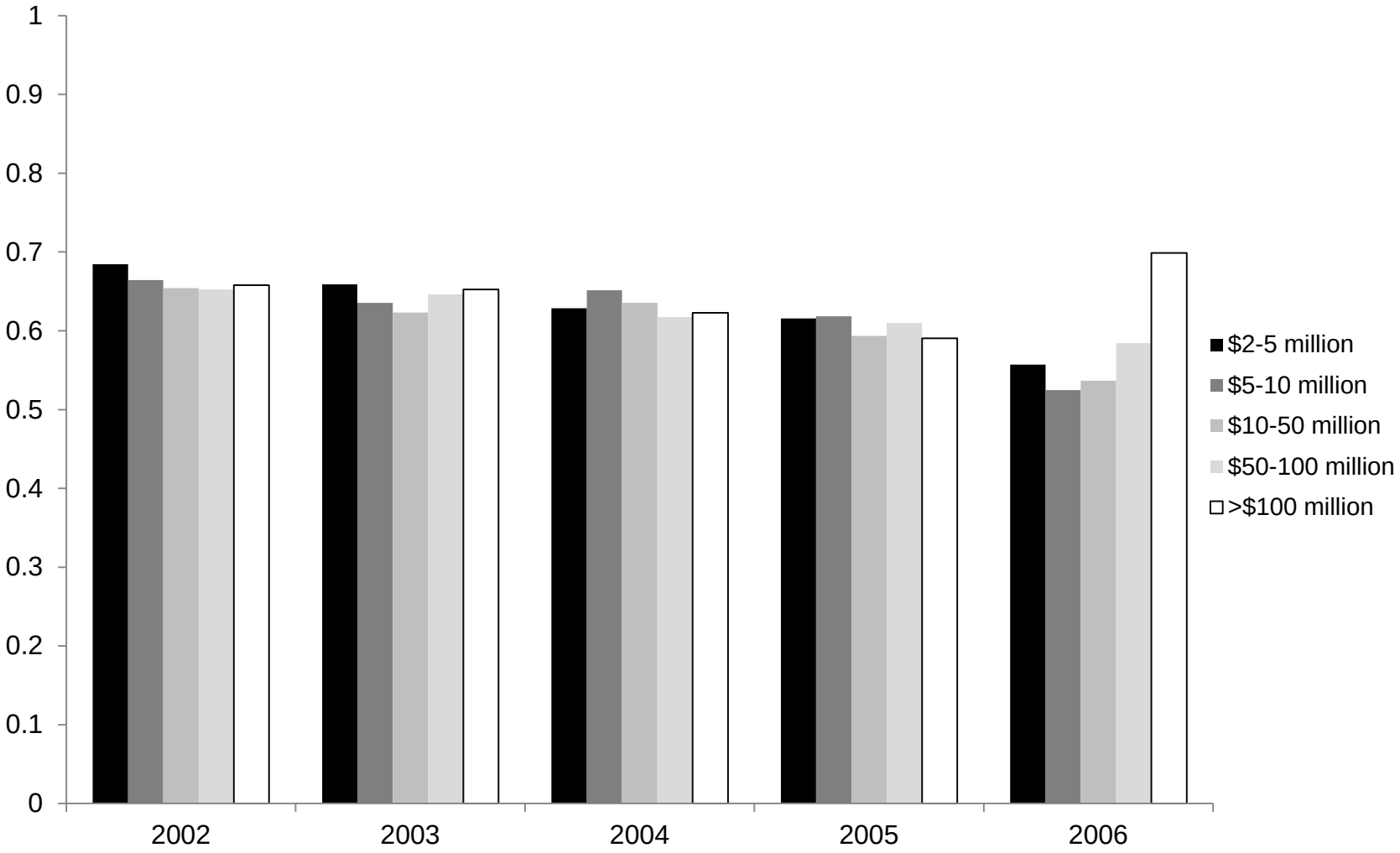




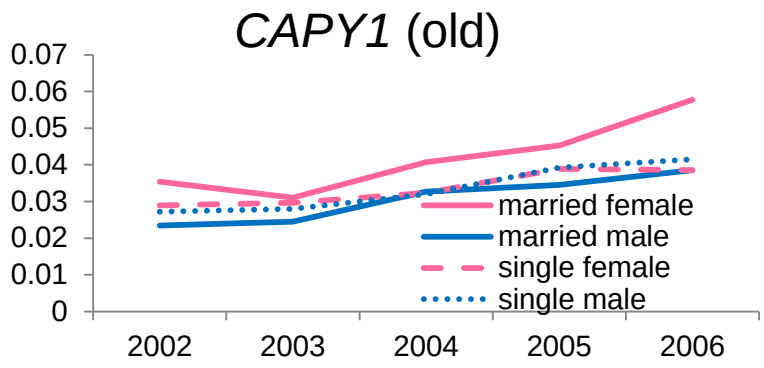
Capital Gains as a Proportion of CAPY1, by Wealth Category (2002-2006)

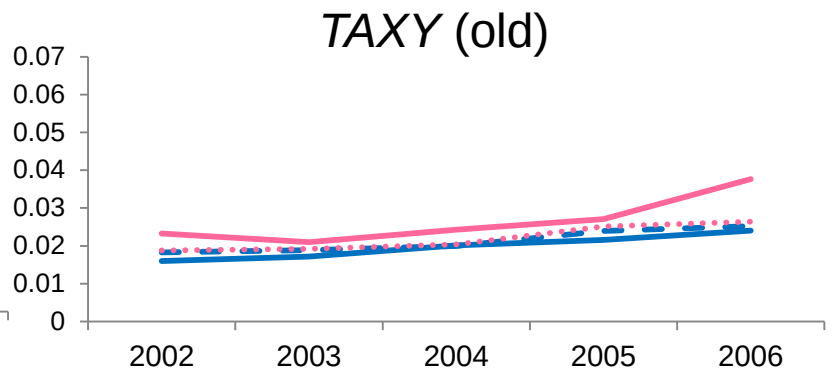


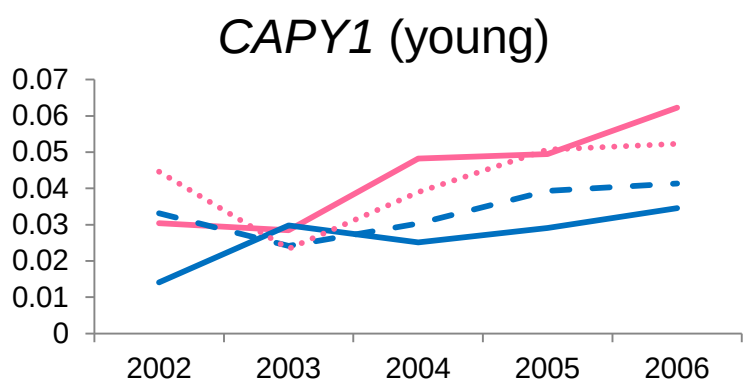
TAXY as a Proportion of CAPY1, by Wealth Category (2002-2006)

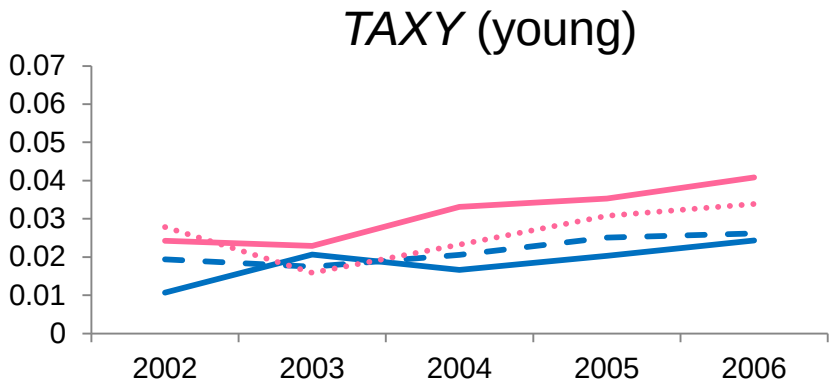


Net Capital Income as Proportion of Net Estate (2002-2006)

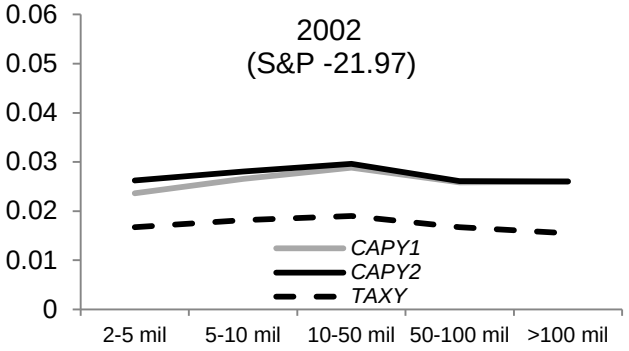


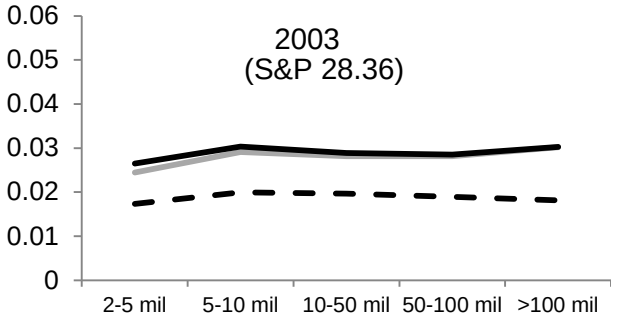


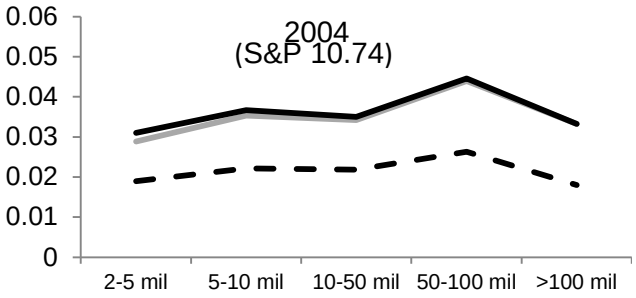


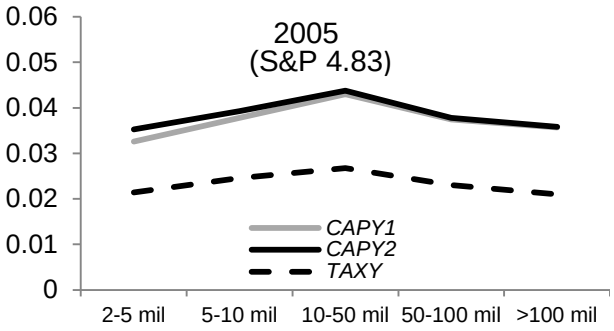


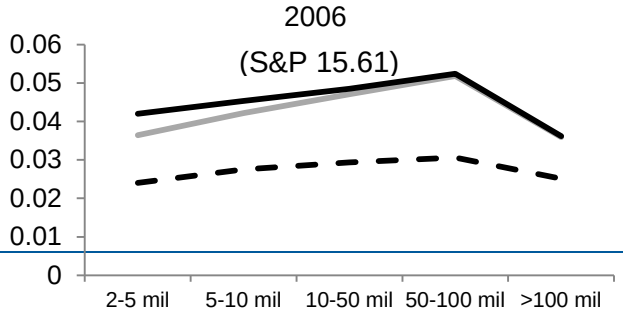
Measures of Net Return to Capital by Net Estate Category, Separately by Year



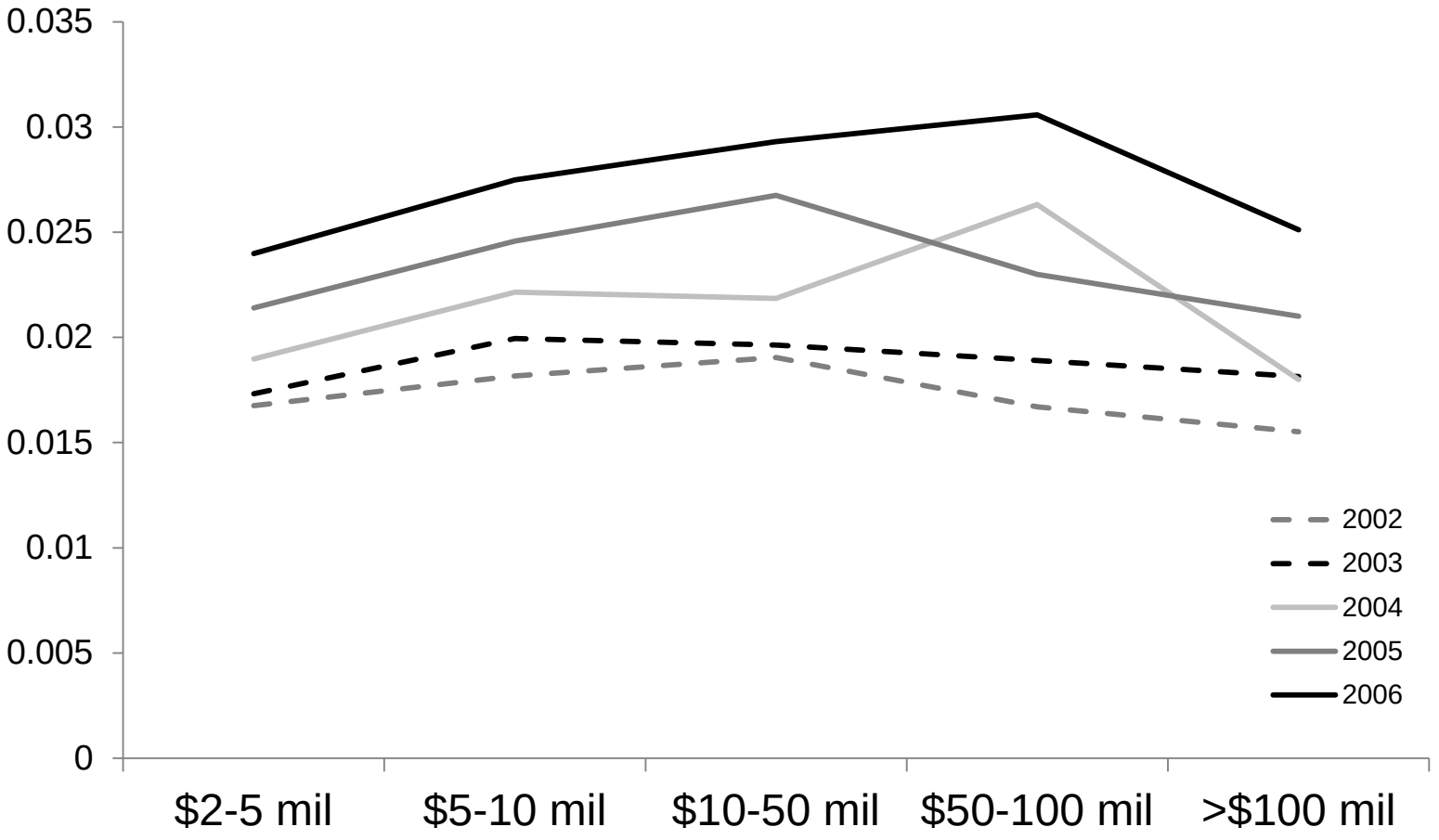




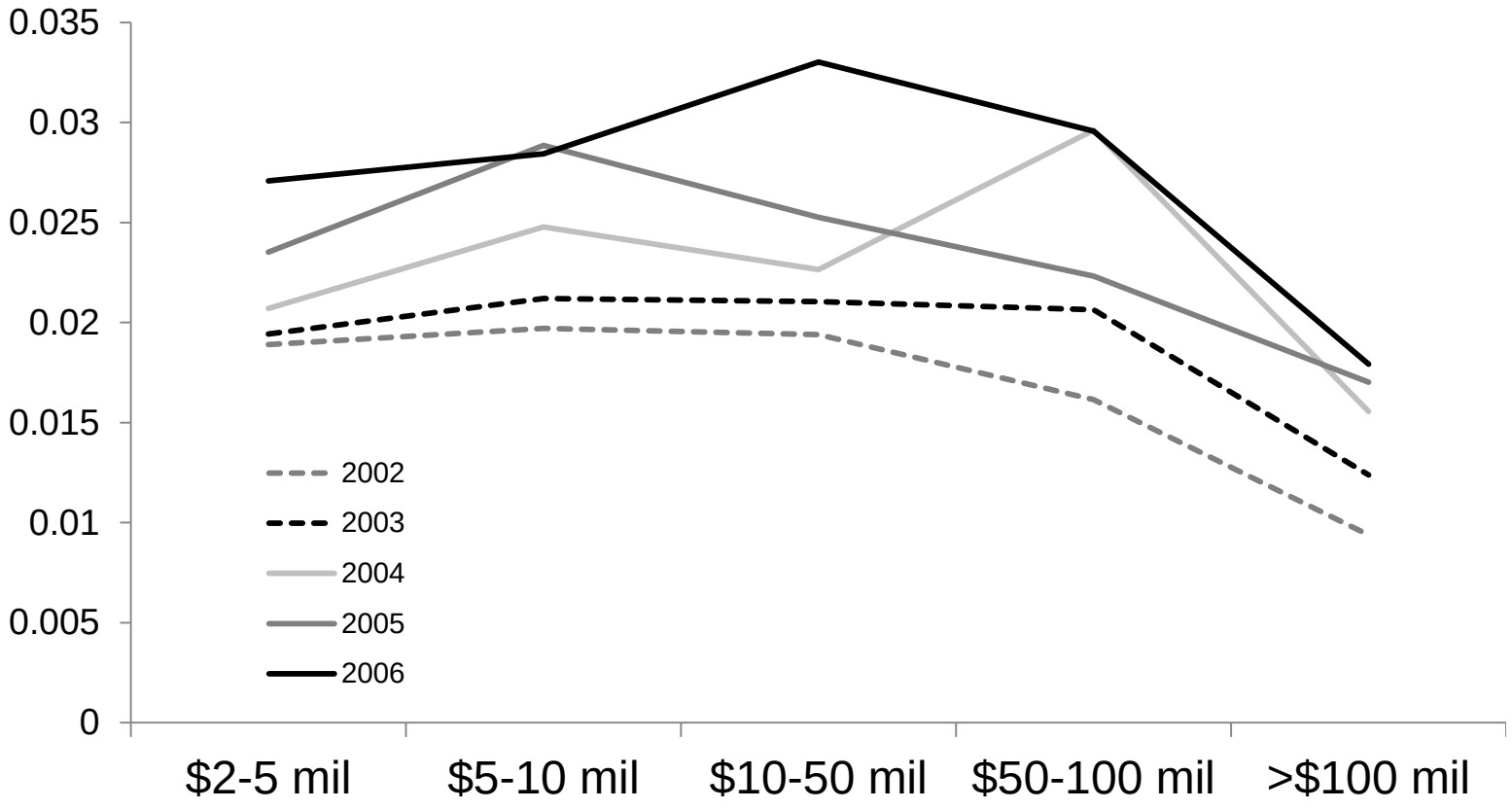




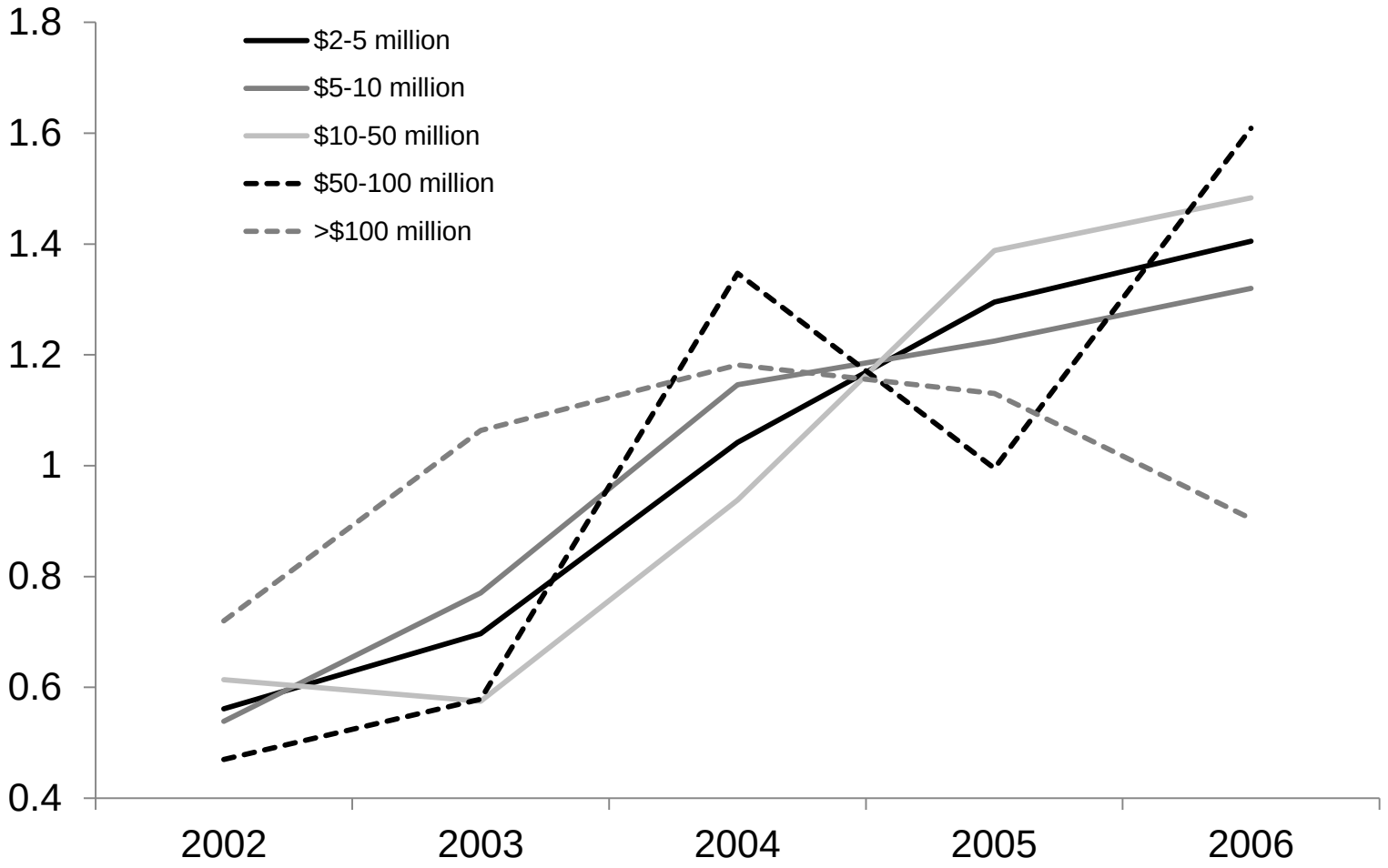
Taxable Capital Income as a Percentage of Net Estate by Wealth Category



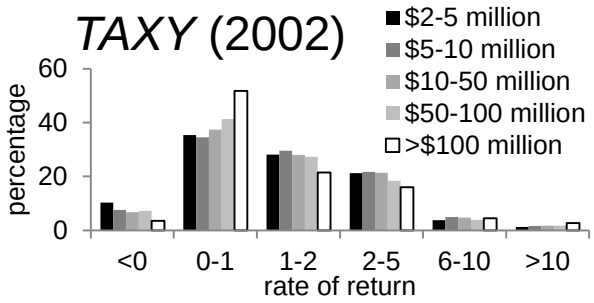
Taxable Capital Income as a Percentage of Net Estate by Wealth Category (non-homeowners)

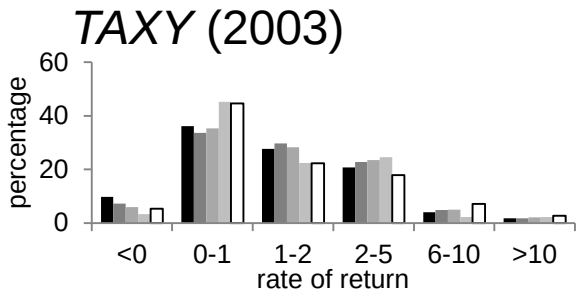


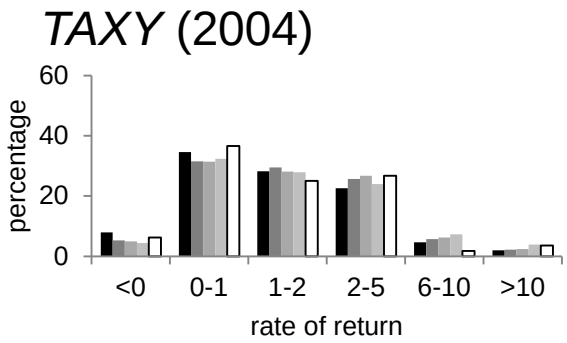
Proportion of Realized Capital Gains Relative to 5-Year Average, by Wealth Class, 2002--2006

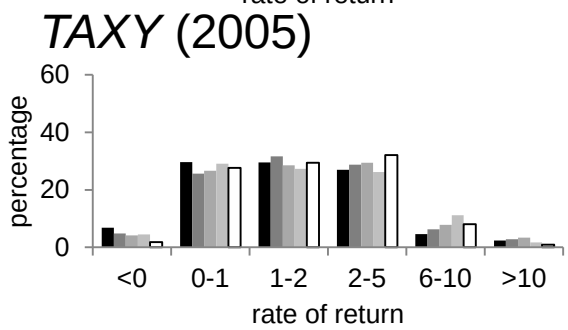


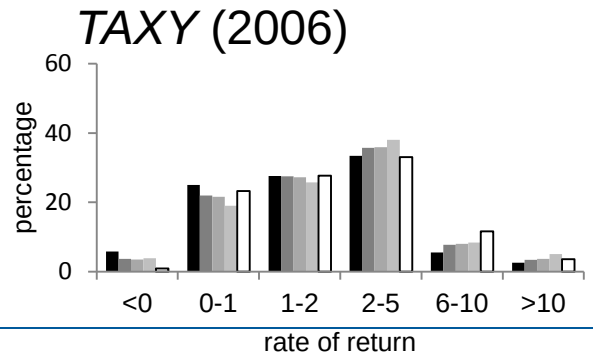
Percentages of Estates with Taxable Capital Income in a Particular Range, by Wealth Category











Regression Analysis

$\ln(\text{TAXY}) = \alpha + \beta_1 \ln \text{net estate} + \beta_2 \text{age} + \beta_3 \text{age squared} +$

β_4 (D always married) + β_4(D always single) + β_5(D male) +

β_6(D male*always married) + β_7(D male*always single) +

$\beta_8 \ln$ charitable deduction + β_9 homepct +

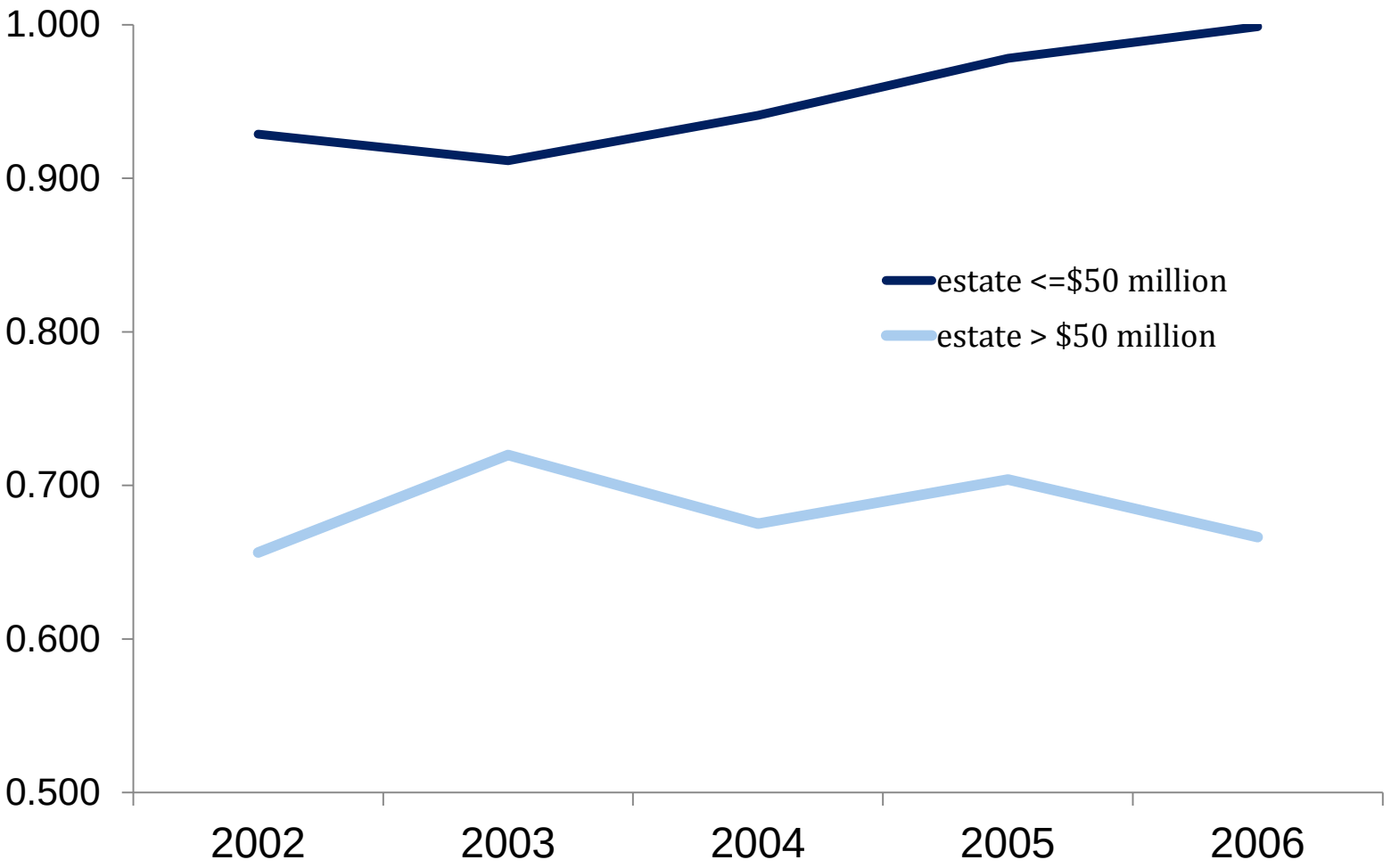
β_{10}(D estate>$50 million) +

β_{11} (D estate>$50 million * ln net estate)

Robust standard errors

Adjusted R-squared 0.304-0.340

Elasticity of Taxable Capital Income with respect to Net Estate, by Year and Wealth Category



Tax Policy Implications

Stock market gains 2003 = 28.36%

Long-term real return on stock ~ 7%

Typical realized return by wealthy individual ~ 2%

Top marginal income tax rate = 35%

Effective marginal income tax rate = 10%

Caveat: does not account for other taxes paid

The Wisdom of Warren Buffett

"I still pay a lower tax rate than my secretary"

Compares capital gains tax rate to rate on labor income

Doesn't account for discretion in realization

Wealthy have higher economic returns and lower realized returns than non-wealthy

Inferring wealth distribution from realized capital income understates wealth inequality

Measuring Income and Wealth at the Top Using Administrative and Survey Data

Jesse Bricker

Alice Henriques

Jacob Krimmel

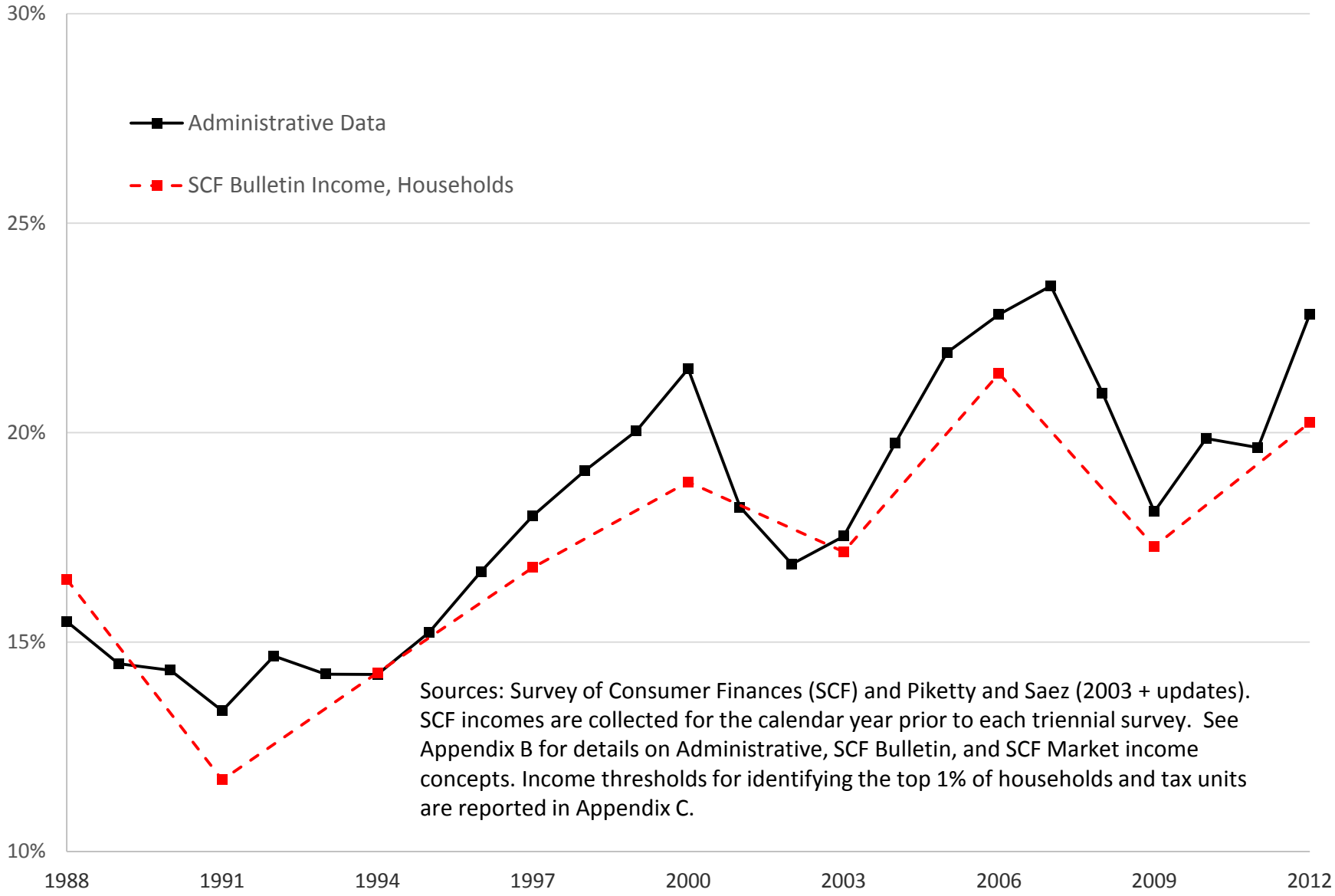
John Sabelhaus

Presentation prepared for SOI Advisory Panel Meeting, June 5, 2015. The analysis and conclusions set forth are those of the author and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve System.

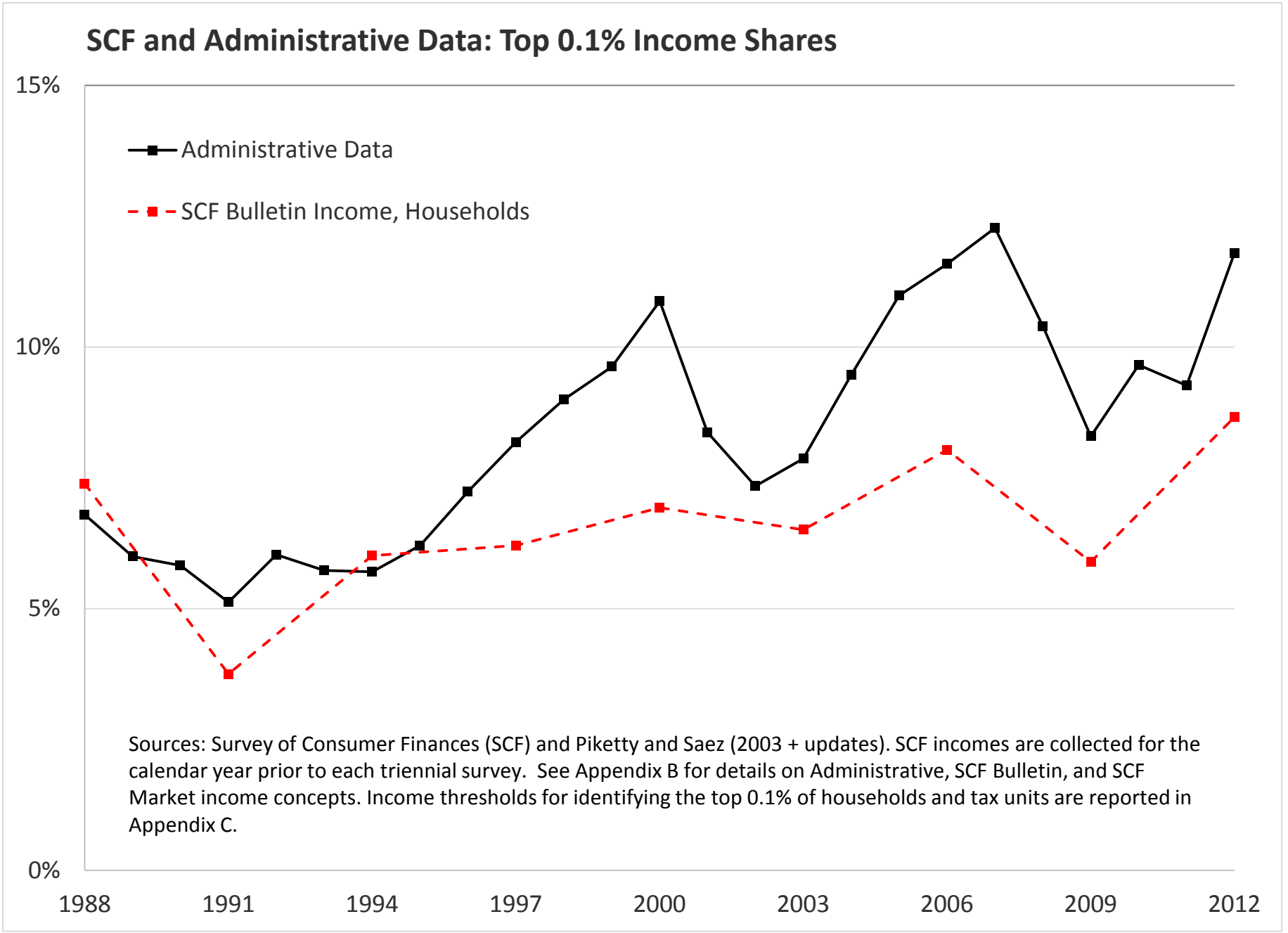
Goals for this Paper

- U.S. top income and wealth shares are high and rising, but how high, and how fast?
- Widely-cited top shares estimates based on administrative income tax data diverge from Survey of Consumer Finances (SCF)
  - Piketty and Saez (2003, updated)
  - Saez and Zucman (2014)
- Primary goal is to understand *why* the two approaches diverge, and solve for biases

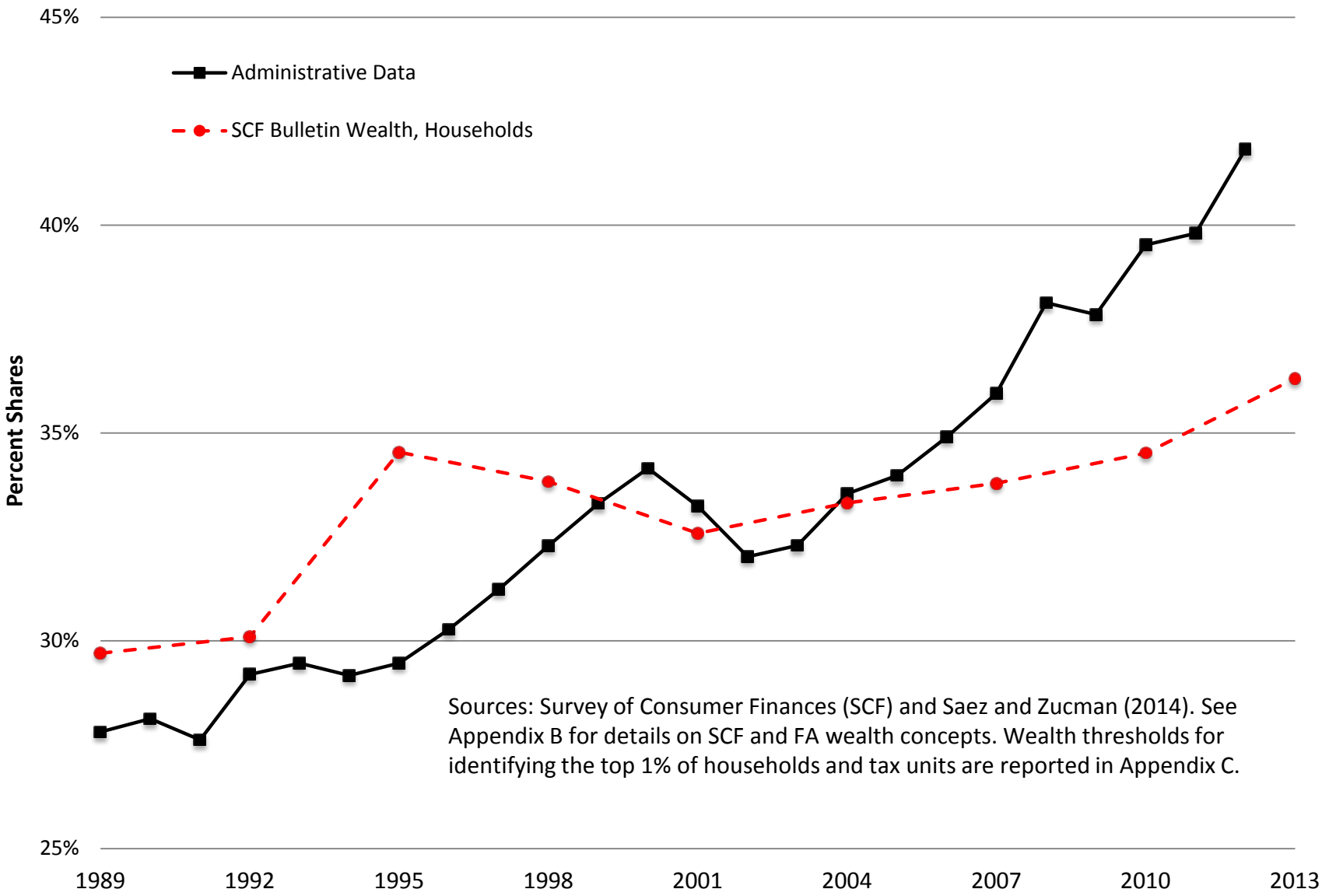
SCF and Administrative Data: Top 1% Income Shares



Sources: Survey of Consumer Finances (SCF) and Piketty and Saez (2003 + updates). SCF incomes are collected for the calendar year prior to each triennial survey. See Appendix B for details on Administrative, SCF Bulletin, and SCF Market income concepts. Income thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

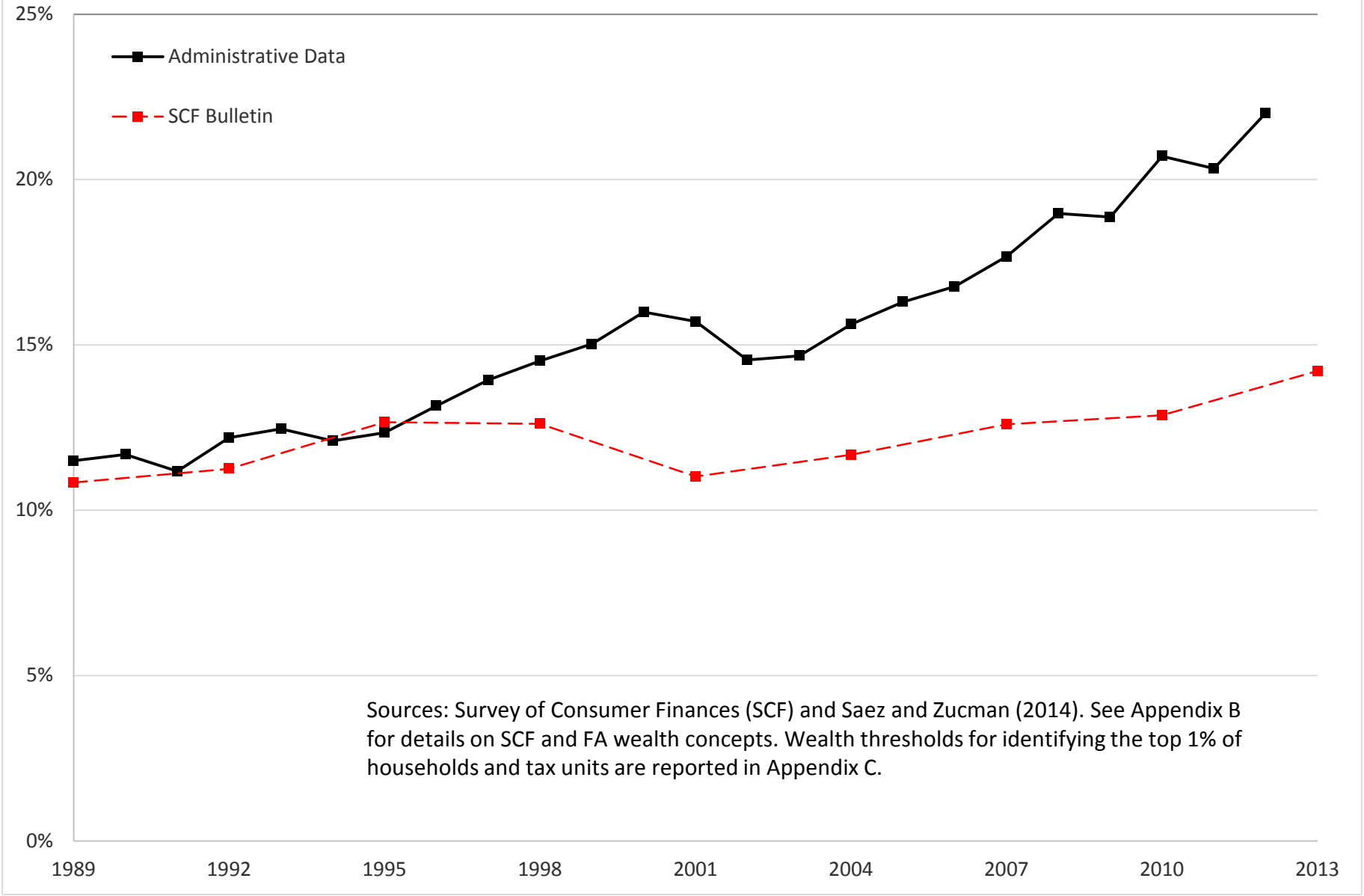

SCF and Administrative Data: Top 0.1% Income Shares
15%
10%
5%
0%
Administrative Data
SCF Bulletin Income, Households
1988
1991
1994
1997
2000
2003
2006
2009
2012
Sources: Survey of Consumer Finances (SCF) and Piketty and Saez (2003 + updates). SCF incomes are collected for the calendar year prior to each triennial survey. See Appendix B for details on Administrative, SCF Bulletin, and SCF Market income concepts. Income thresholds for identifying the top 0.1% of households and tax units are reported in Appendix C.

SCF and Administrative Data Top 1% Wealth Shares



Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

SCF and Administrative Data: Top 0.1% Wealth Shares



Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

Top Wealth Shares Reconciliation

- Why do SCF and tax-based "Gross Capitalization" top wealth share estimates diverge?
  - Capitalized approach uses taxable SOI incomes for income-generating assets, imputations for rest
  - Capitalized calibrated to Financial Accounts (FA)
  - SCF and FA balance sheet *concepts* diverge
  - SCF and FA estimated *aggregates* diverge
  - Some implied capitalization factors problematic
- 160 million tax units versus 120 million families
- SCF (by rule) does not survey Forbes 400

Gross Capitalization (GC) Approach

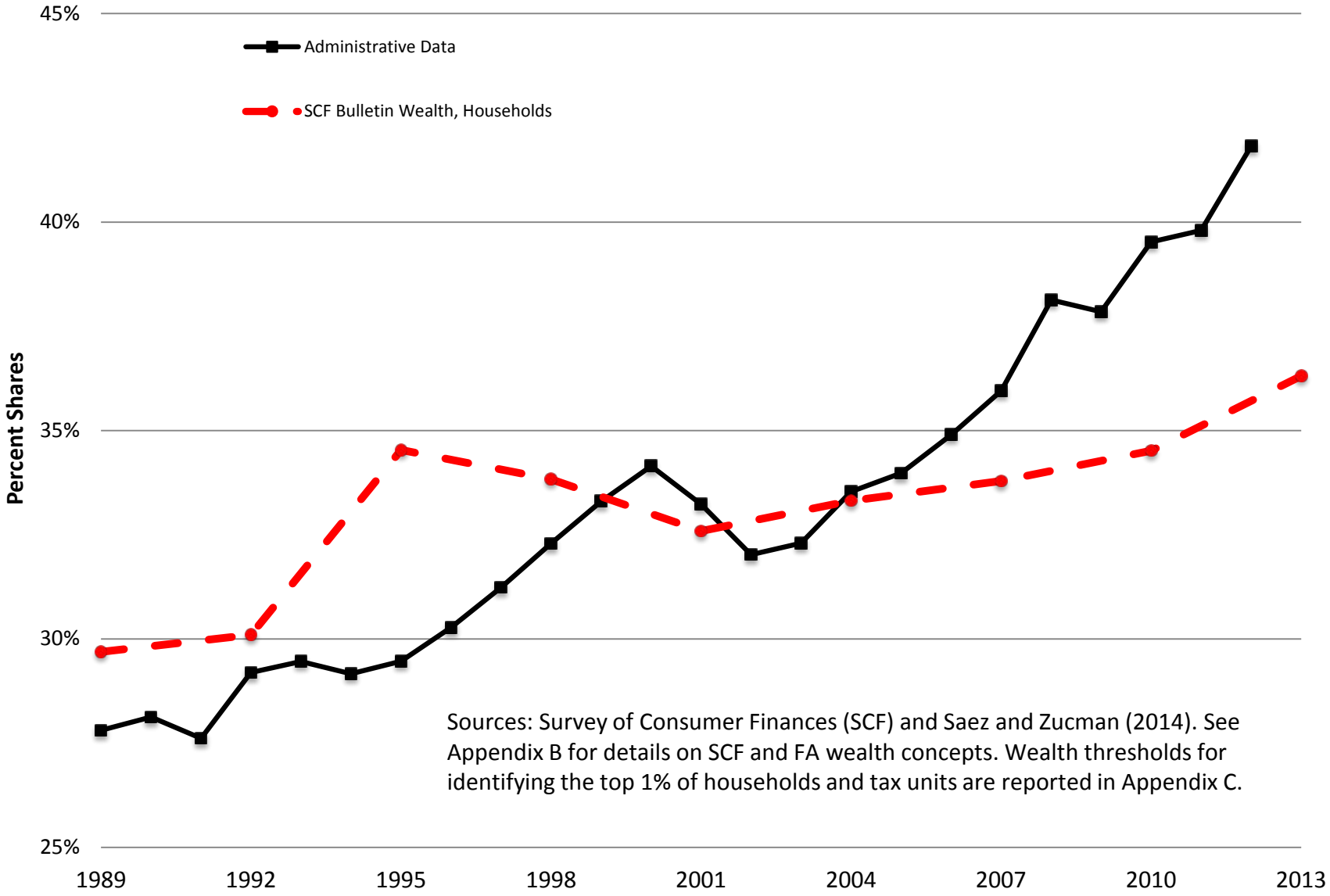
- Given taxable capital income type *k*=1,…9 along with estimates of wealth that do not generate taxable income, for family *i*

$$\widehat{wealth}_i^{GC} = \sum_{\forall k} \frac{SOI\ income_i^k}{ror^k} + nonfinancial_i$$

- In practice, Saez and Zucman (2014) compute *ror* for each asset *k* to calibrate to FA aggregates

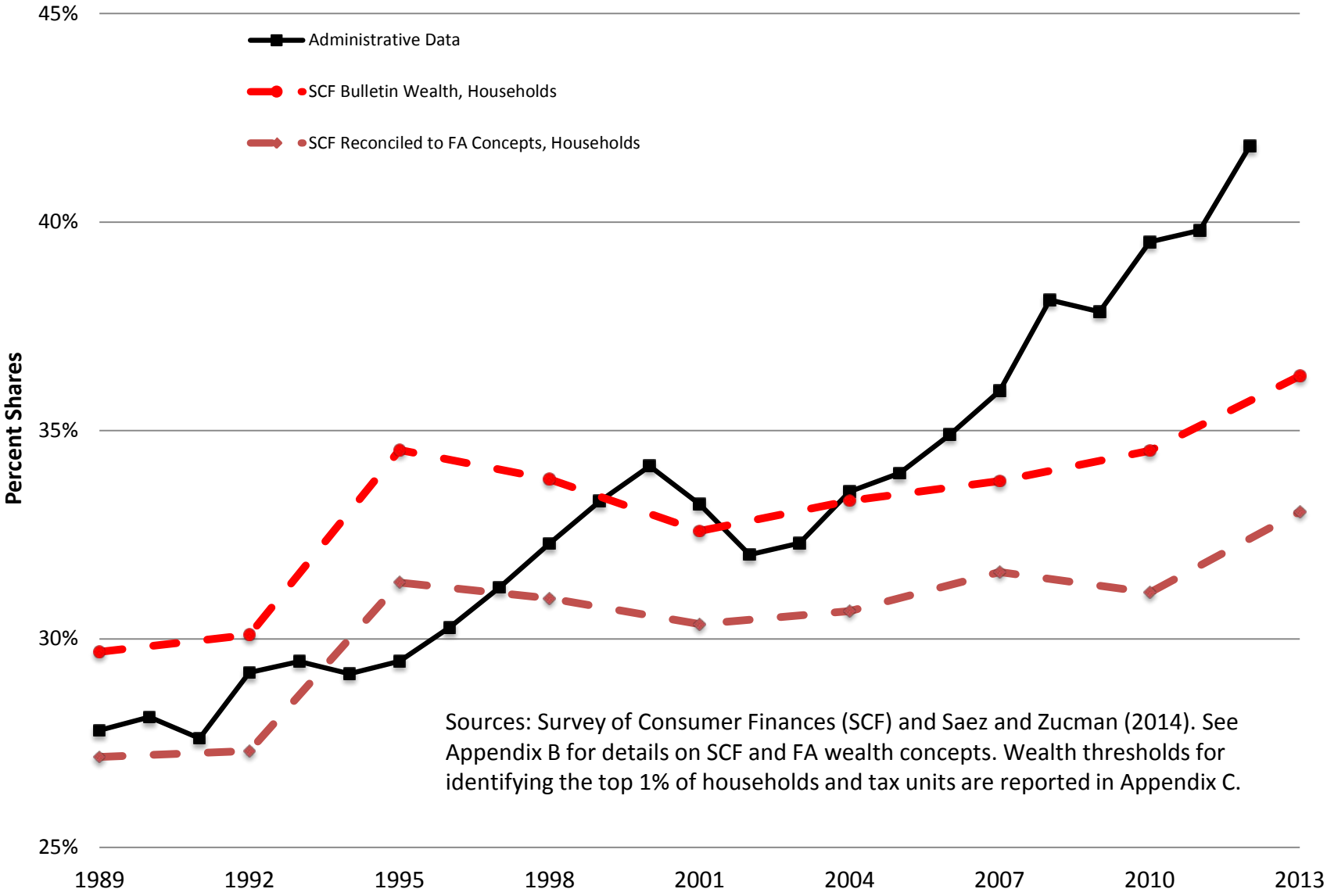
$$ror^k = \frac{\sum_{\forall i} SOI\ income_i^k}{FA\ asset^k}$$

Reconciling Survey of Consumer Finances (SCF) and Administrative Data
Top 1% Wealth Shares



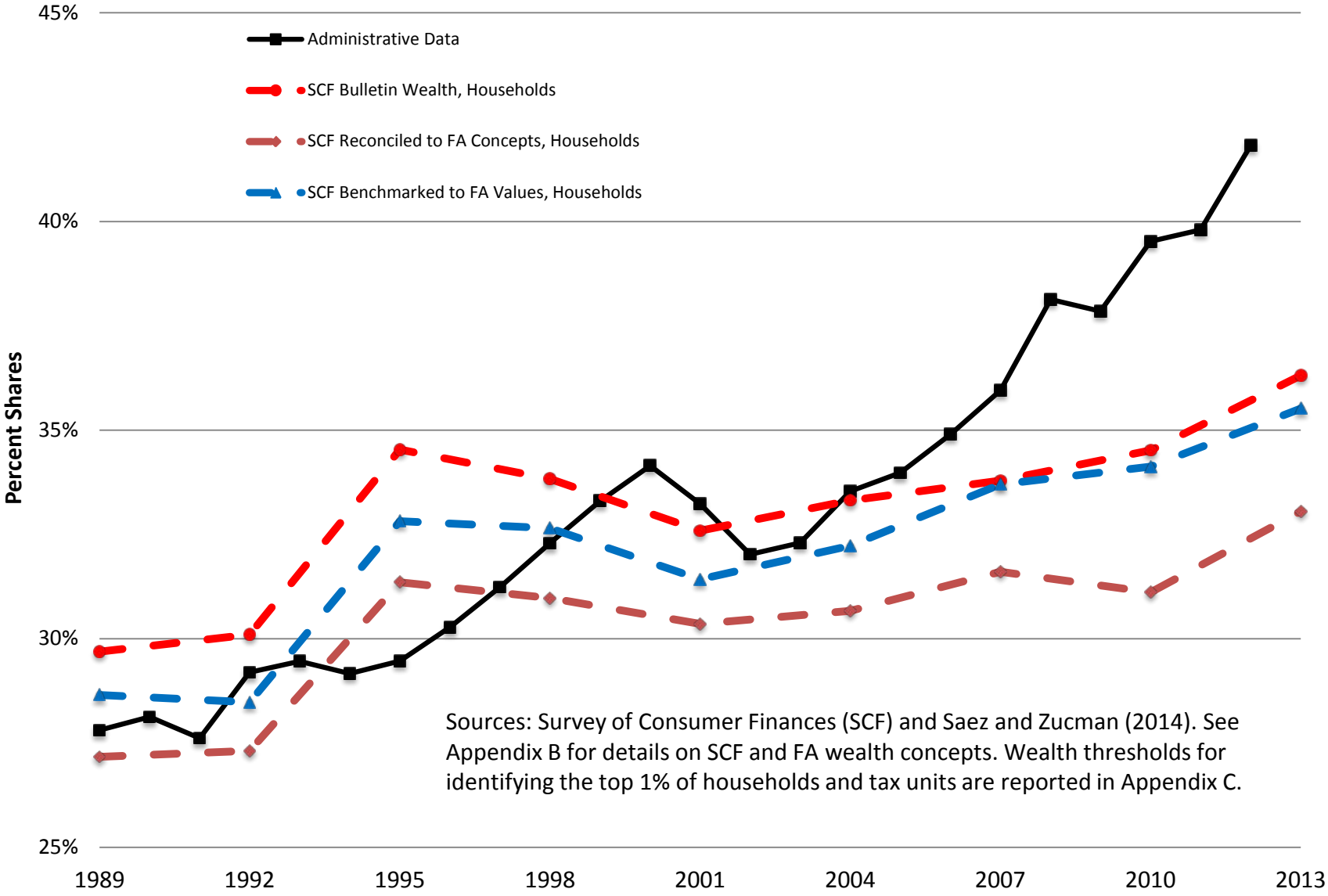
Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

Reconciling Survey of Consumer Finances (SCF) and Administrative Data
Top 1% Wealth Shares



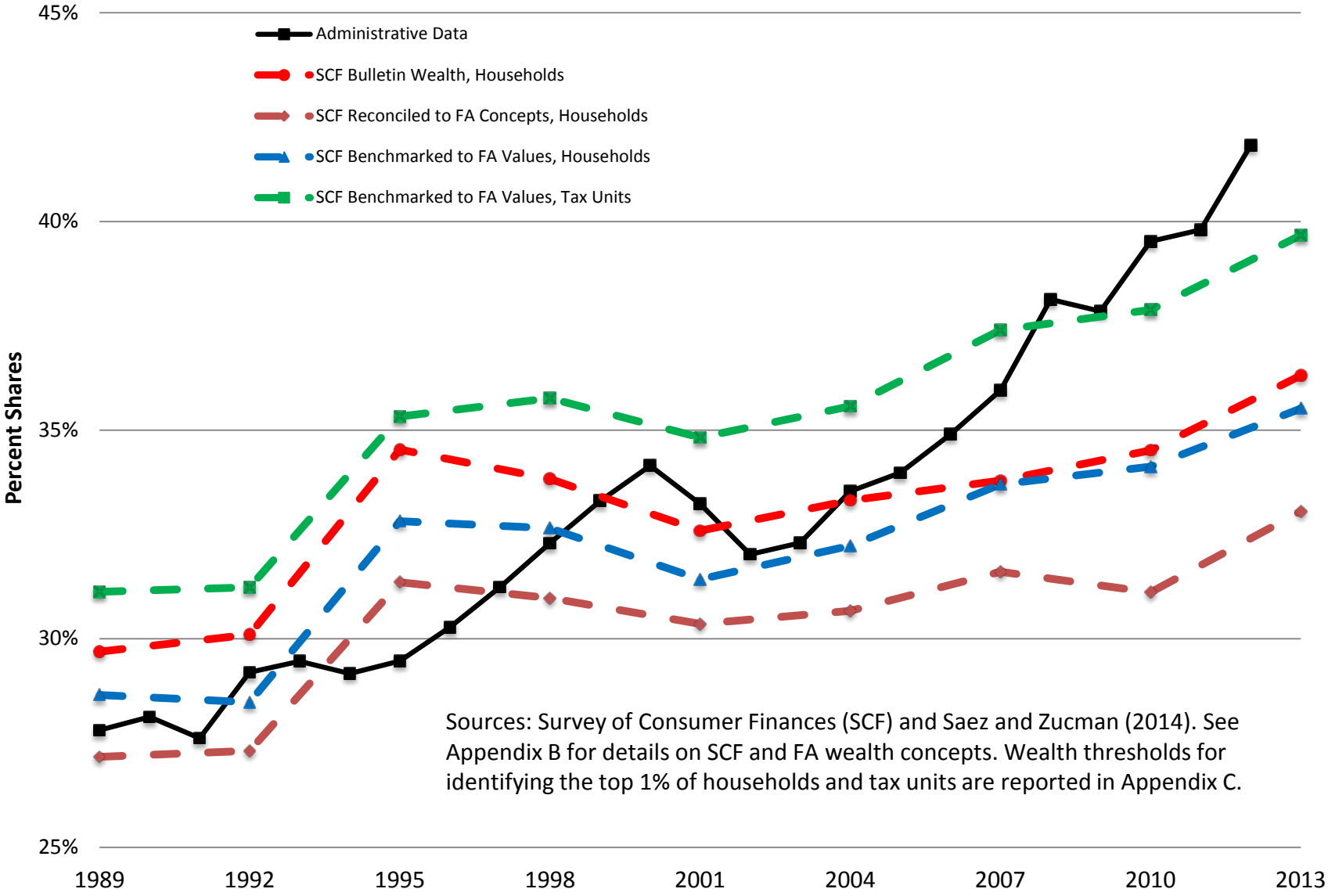
Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

Reconciling Survey of Consumer Finances (SCF) and Administrative Data
Top 1% Wealth Shares



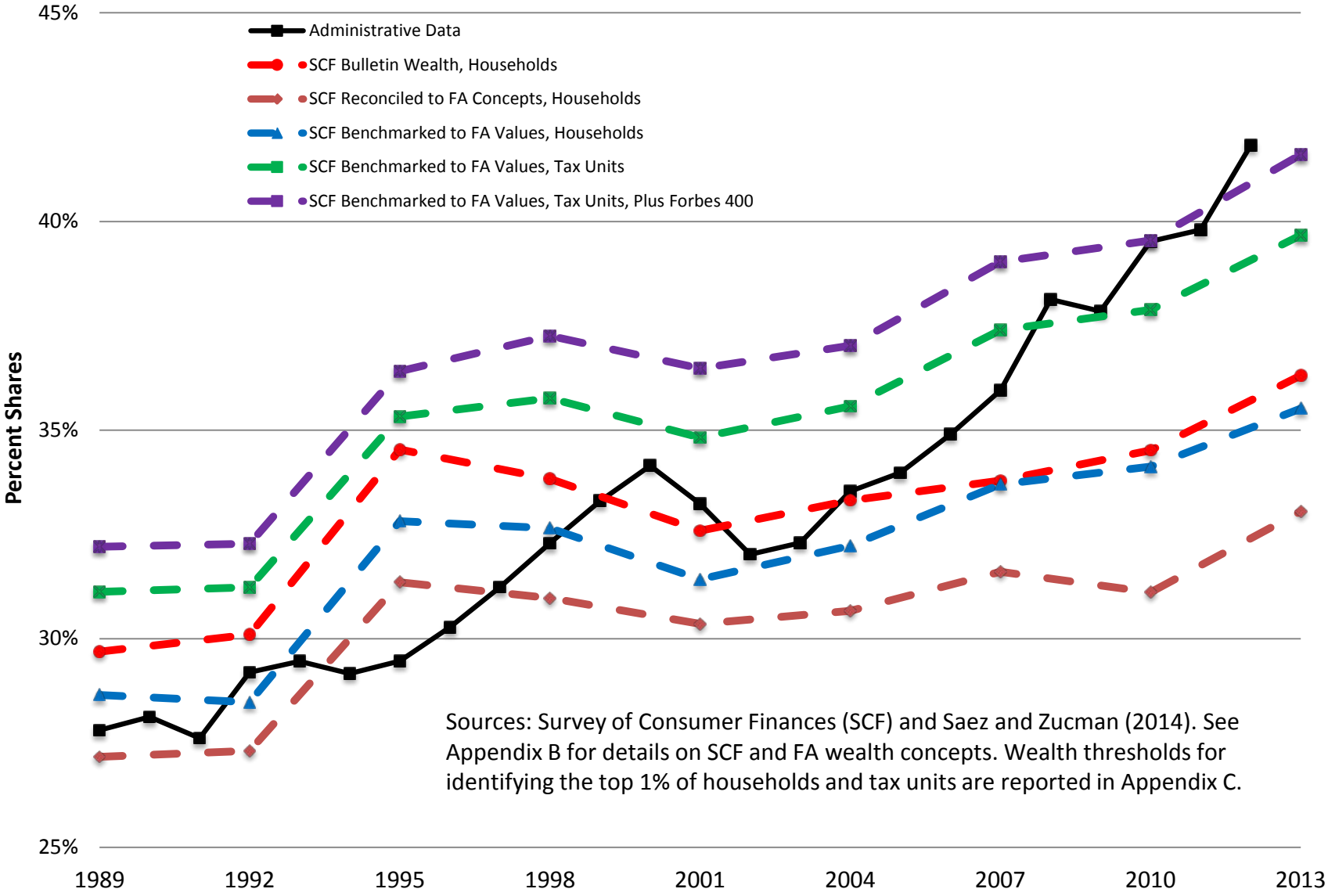
Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

Reconciling Survey of Consumer Finances (SCF) and Administrative Data
Top 1% Wealth Shares

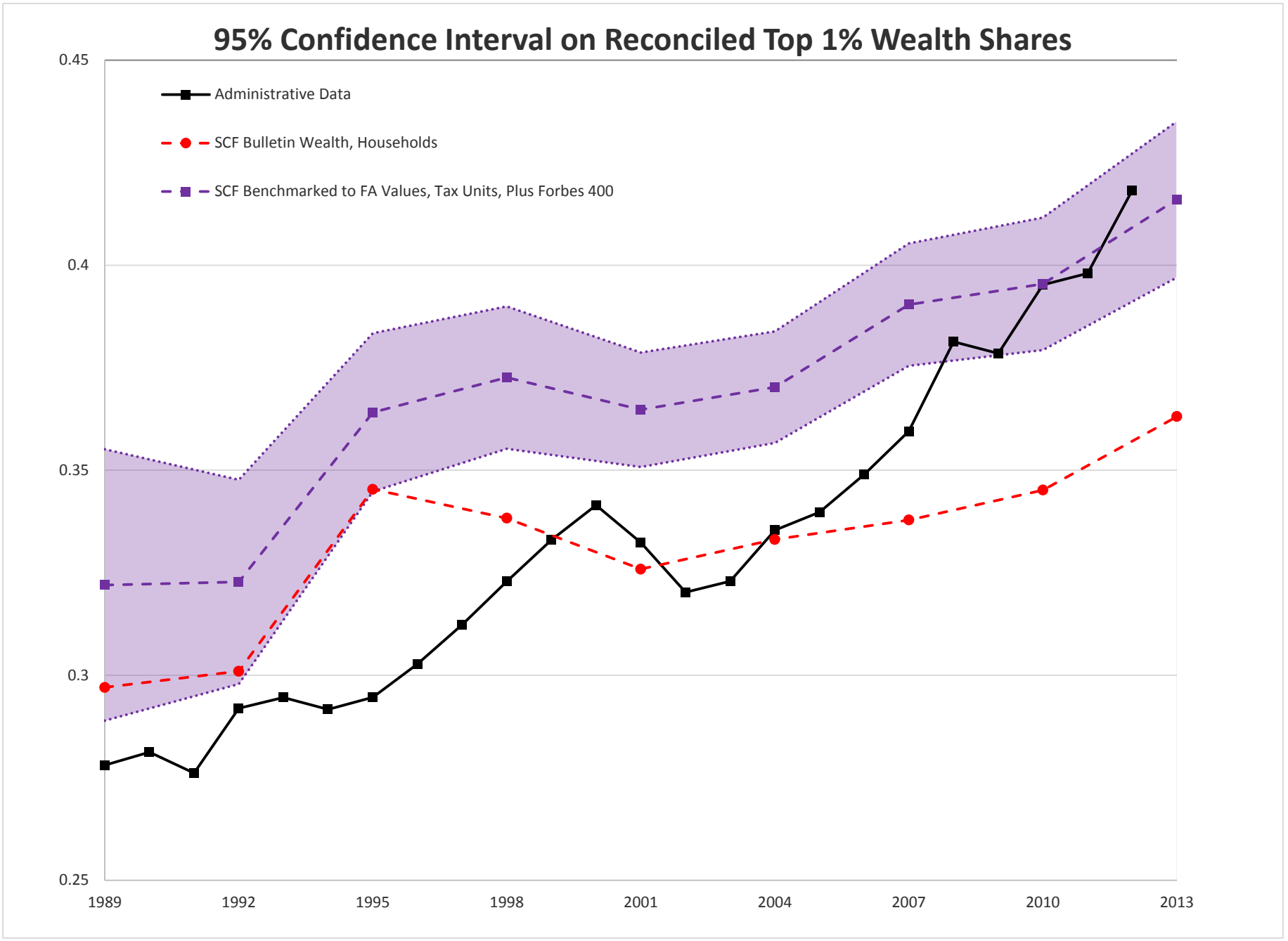


Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.

Reconciling Survey of Consumer Finances (SCF) and Administrative Data
Top 1% Wealth Shares

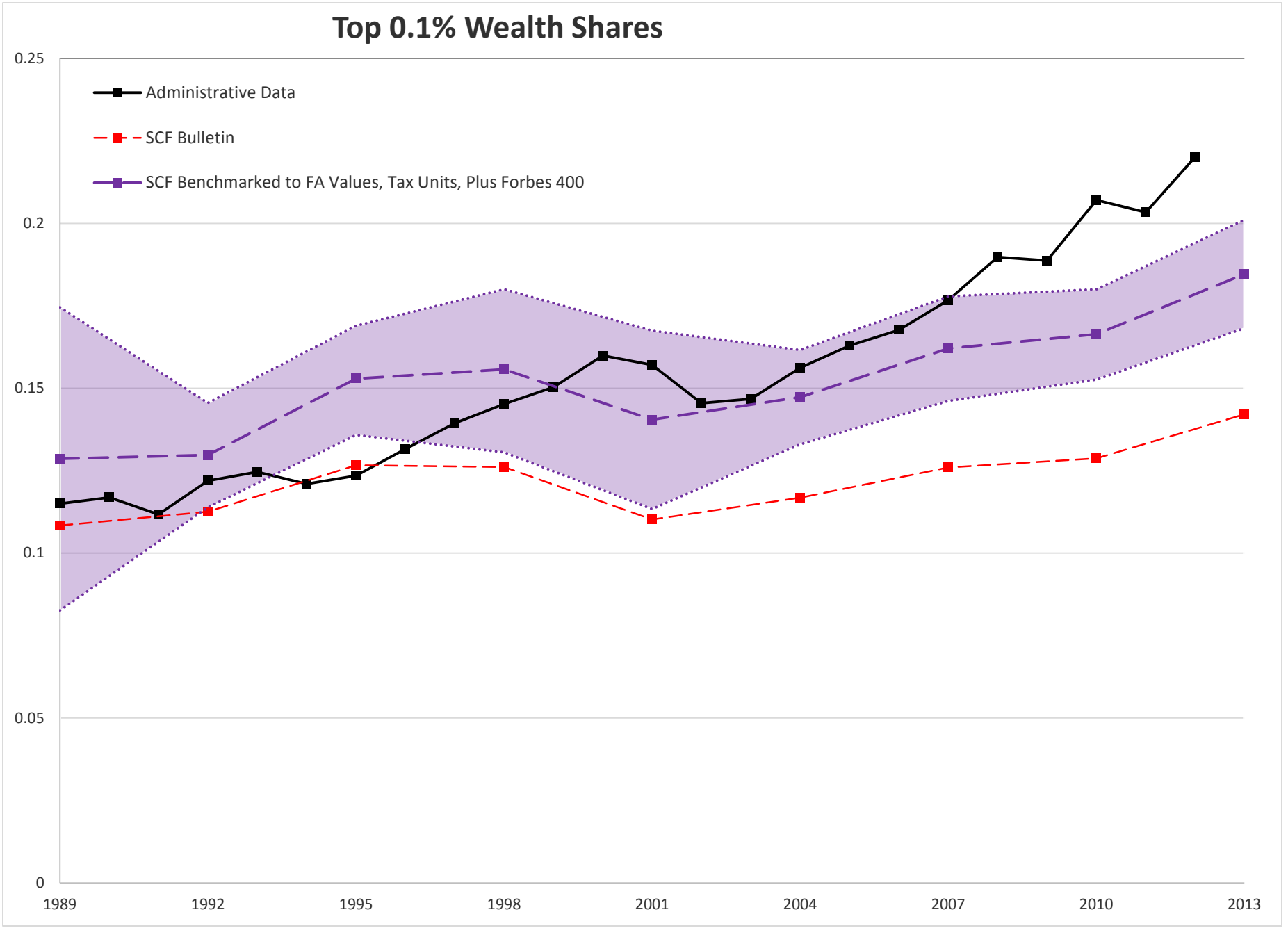


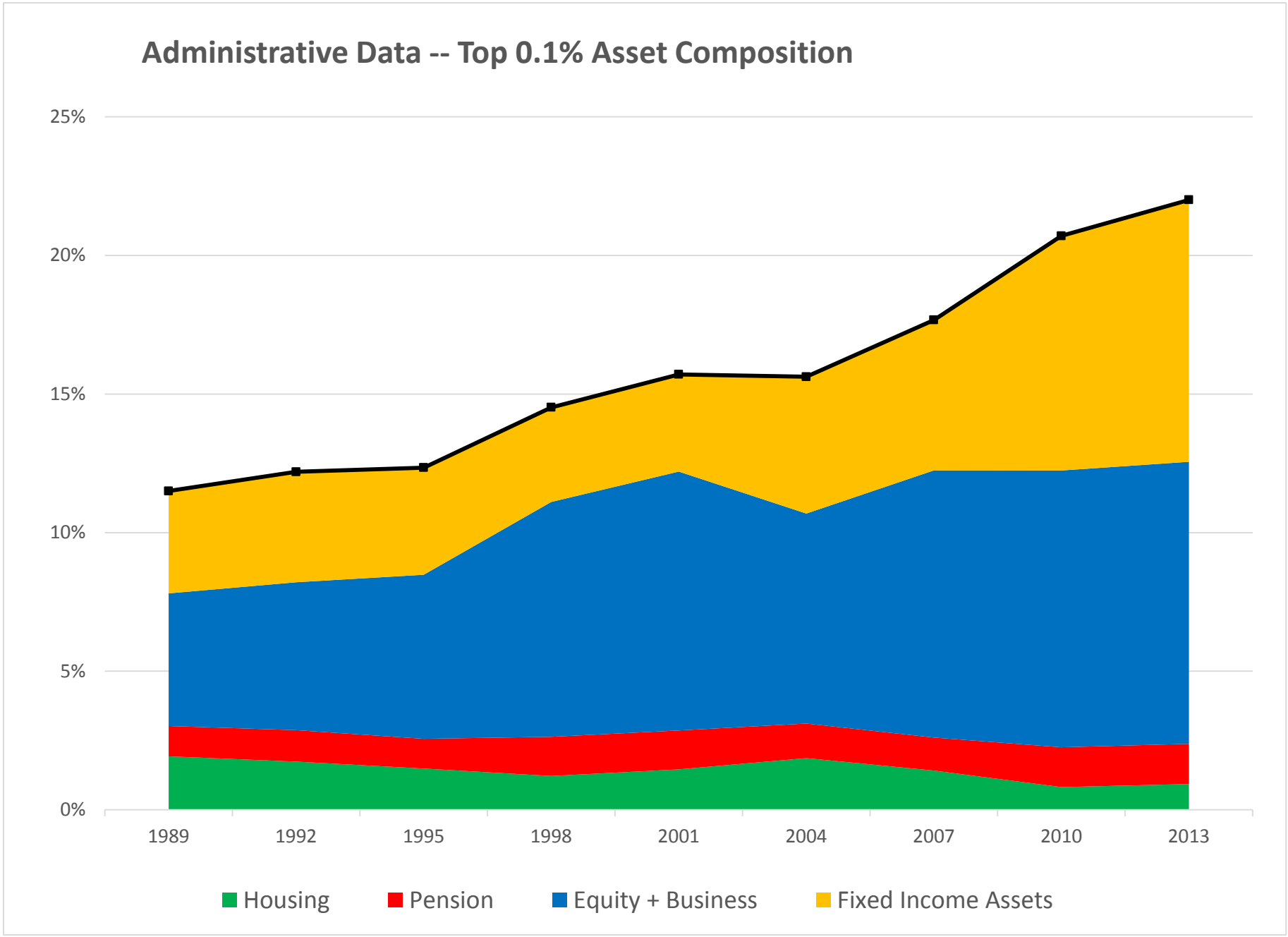
Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C.


95% Confidence Interval on Reconciled Top 1% Wealth Shares
0.45
0.4
0.35
0.3
0.25
Administrative Data
SCF Bulletin Wealth, Households
SCF Benchmarked to FA Values, Tax Units, Plus Forbes 400
1989
1992
1995
1998
2001
2004
2007
2010
2013

Wealth Reconciliation at the Very Top

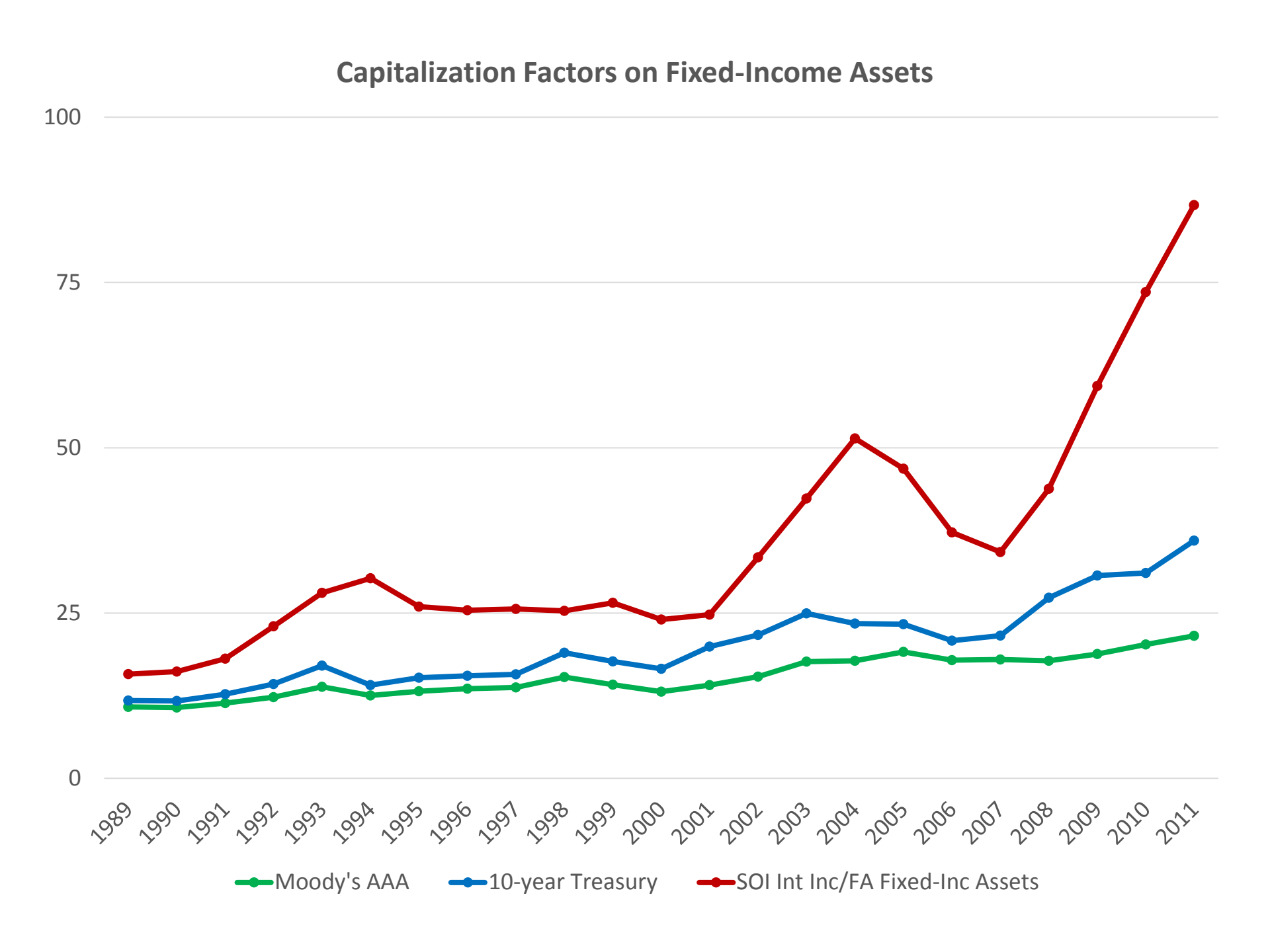
- Recent top 1% wealth shares largely reconciled, some remaining trend divergence
- Still, recent top 0.1% wealth share is greater in capitalized administrative tax data

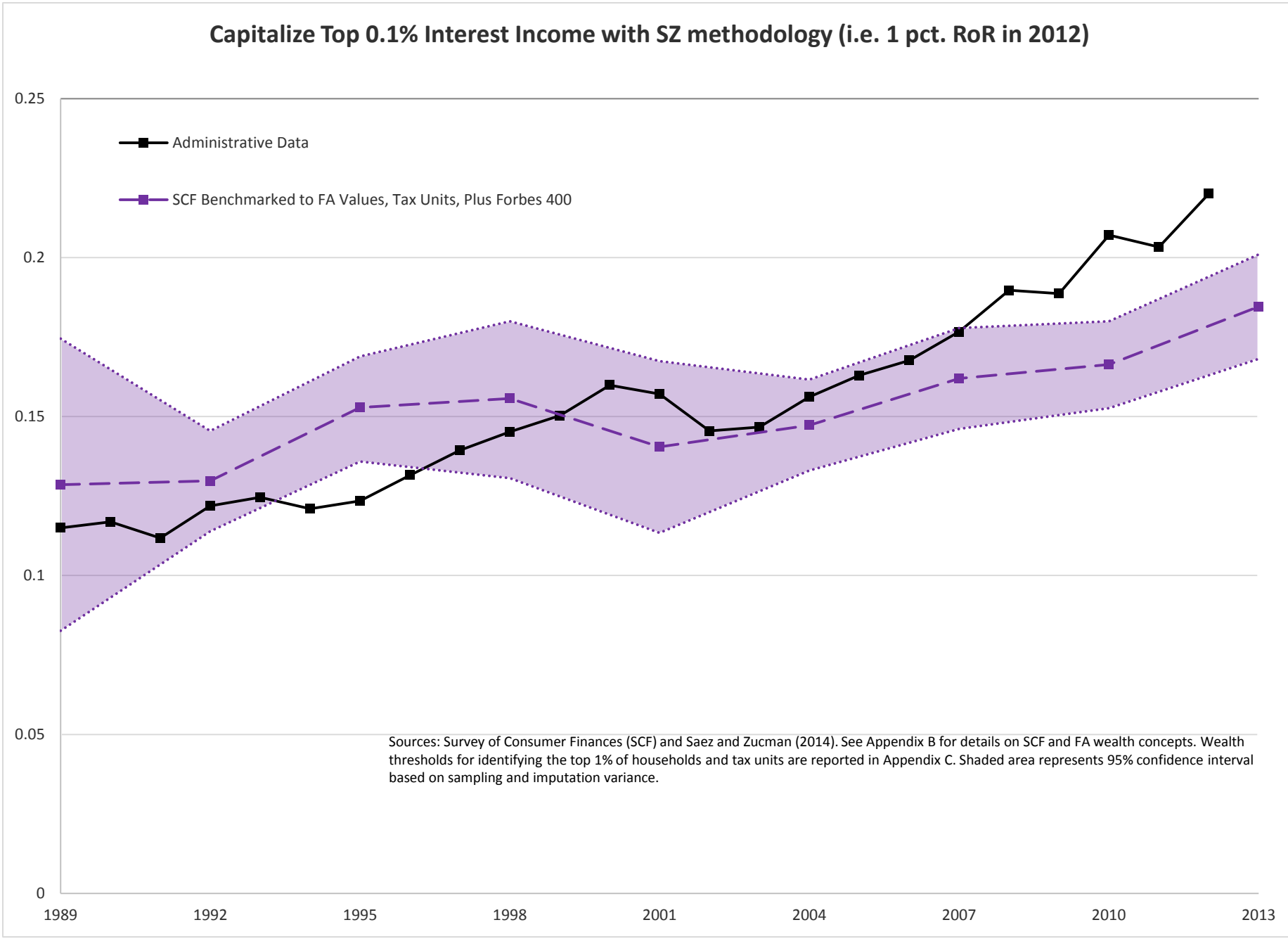

Top 0.1% Wealth Shares
Administrative Data
SCF Bulletin
SCF Benchmarked to FA Values, Tax Units, Plus Forbes 400
0.25
0.2
0.15
0.1
0.05
0
1989
1992
1995
1998
2001
2004
2007
2010
2013

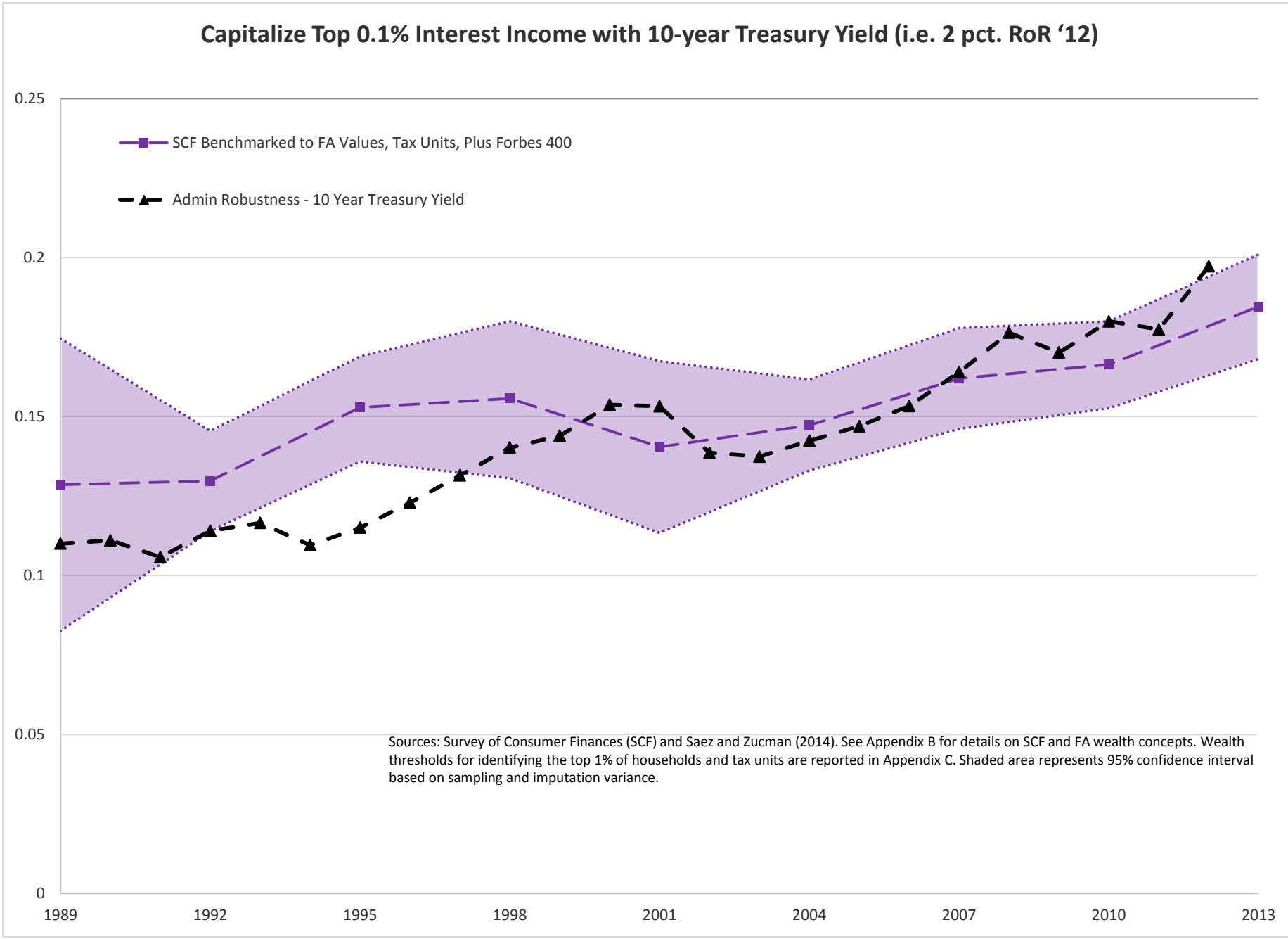

Administrative Data -- Top 0.1% Asset Composition
25%
20%
15%
10%
5%
0%
1989
1992
1995
1998
2001
2004
2007
2010
2013
Housing
Pension
Equity + Business
Fixed Income Assets

Wealth Reconciliation at the Very Top

- Fixed income explains why top 0.1% wealth share greater in capitalized administrative tax data
- Look closer at asset composition and RoR
  - Fixed-income assets were 25%, now 45% of assets
  - Bonds ≈ 1/3rd, deposit accounts are the other 2/3^{rds}.
  - Do the top 0.1 really hold savings deposit accounts?
- Rate of return on fixed-income = 1 pct. (for all)
  - → capitalization factor of 100x for interest income
  - Compare to market rates of return


Capitalization Factors on Fixed-Income Assets
100
75
50
25
0
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
Moody's AAA
10-year Treasury
SOI Int Inc/FA Fixed-Inc Assets


Capitalize Top 0.1% Interest Income with SZ methodology (i.e. 1 pct. RoR in 2012)
0.25
0.2
0.15
0.1
0.05
0
Administrative Data
SCF Benchmarked to FA Values, Tax Units, Plus Forbes 400
1989
1992
1995
1998
2001
2004
2007
2010
2013
Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C. Shaded area represents 95% confidence interval based on sampling and imputation variance.


Capitalize Top 0.1% Interest Income with 10-year Treasury Yield (i.e. 2 pct. RoR '12)
SCF Benchmarked to FA Values, Tax Units, Plus Forbes 400
Admin Robustness - 10 Year Treasury Yield
0.25
0.2
0.15
0.1
0.05
0
1989
1992
1995
1998
2001
2004
2007
2010
2013
Sources: Survey of Consumer Finances (SCF) and Saez and Zucman (2014). See Appendix B for details on SCF and FA wealth concepts. Wealth thresholds for identifying the top 1% of households and tax units are reported in Appendix C. Shaded area represents 95% confidence interval based on sampling and imputation variance.

Conclusions

- Estimates of top income and wealth shares from SCF can be reconciled with estimates derived directly from administrative tax data
- SCF suggests that administrative-based top share estimates too high and rising too fast
- Reconciliations offer direction for future work, as broader income and wealth measures are likely to further reduce estimated top shares

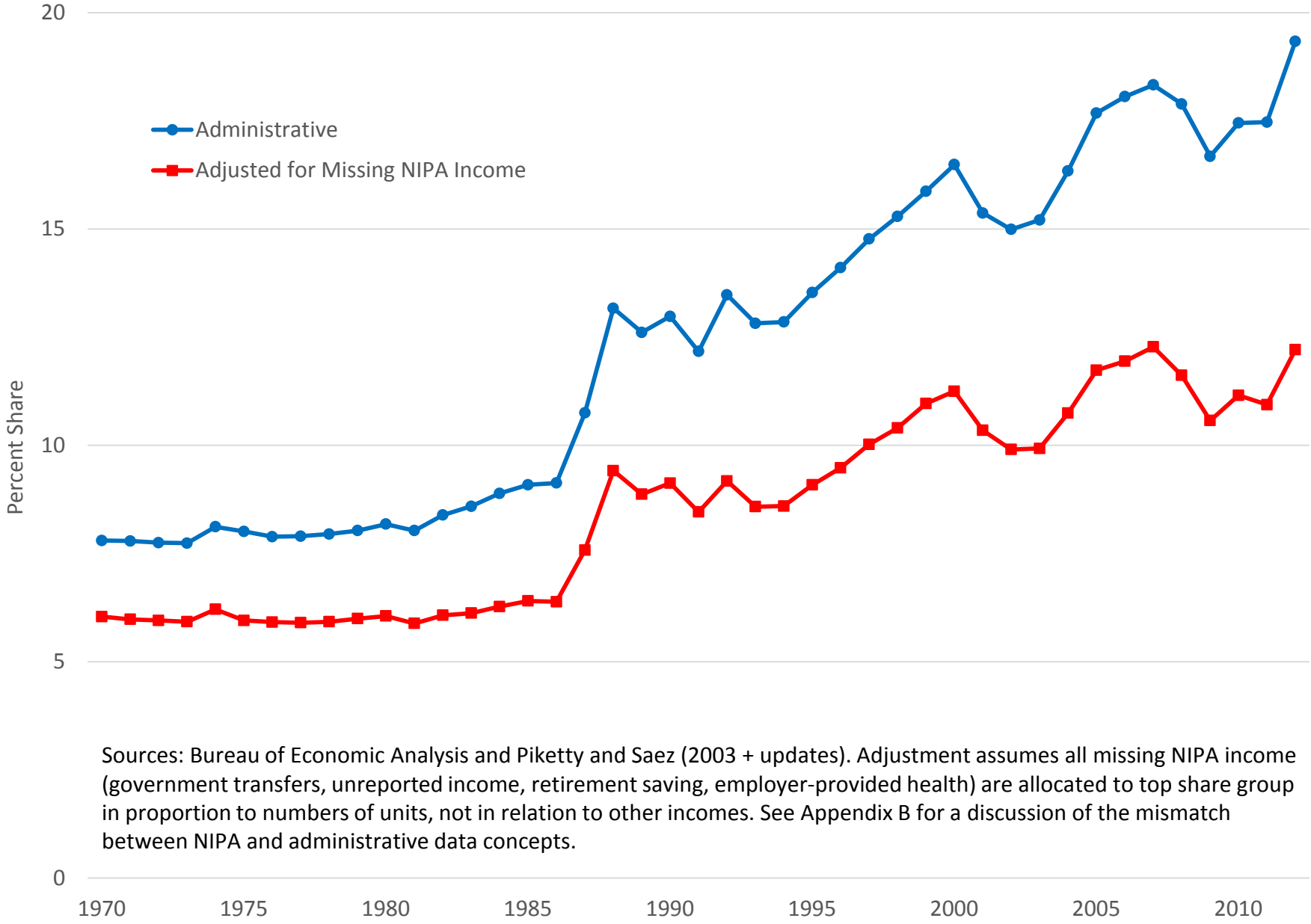
Thanks!

jesse.bricker@frb.gov
alice.henriques@frb.gov
jake.krimmel@frb.gov
john.sabelhaus@frb.gov

If Time: Expanded Income Concept

- Cannot distribute all of NIPA personal income, but can at least bracket top income shares
- Assume that missing income in every year, starting in 1970, is allocated per tax unit
  - Top 1 percent gets only 1% of the missing income
- Top 1% income levels and growth much more muted, and tax unit adjustment would add
- Extreme assumption, but brackets truth: missing incomes are transfers, non-wage compensation, retirement saving

Figure 10. Effect of Allocating Missing Personal Income on Top 1% Income Shares



Sources: Bureau of Economic Analysis and Piketty and Saez (2003 + updates). Adjustment assumes all missing NIPA income (government transfers, unreported income, retirement saving, employer-provided health) are allocated to top share group in proportion to numbers of units, not in relation to other incomes. See Appendix B for a discussion of the mismatch between NIPA and administrative data concepts.

Mortality Differentials – How Much Longevity Can Money Really Buy?

Barry Johnson

Brian Raub

Statistics of Income, IRS

SOI Personal Wealth Study - Background

- Uses Federal estate tax data to estimate wealth of the living population with wealth at or greater than filing threshold
- Based on well-established “Estate Multiplier Technique”

$$MULT = 1 / (p \cdot r) \text{ where:}$$

p = probability of selection to the estate tax sample,

r = mortality rate appropriate to wealthy individuals,

- Assumes that estate tax decedents are random sample of the living wealthy population

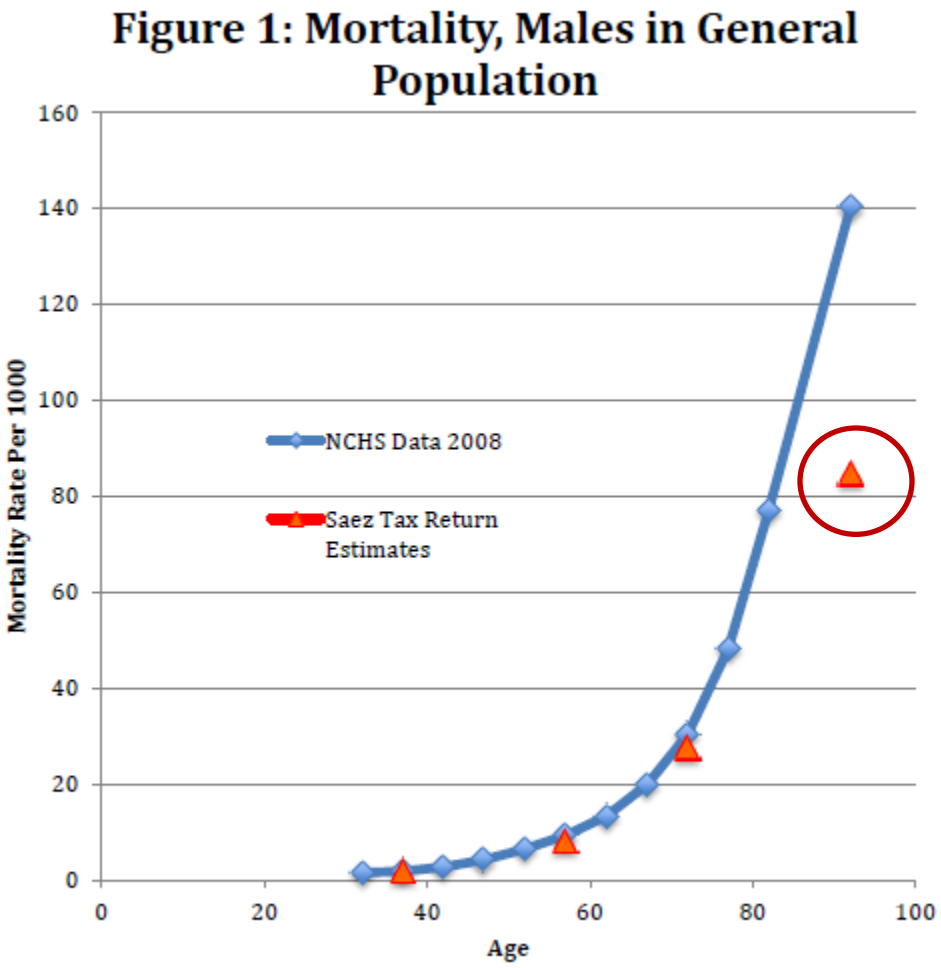
Mortality Rates

- Proper mortality rates for the wealthy are key
  - Estimates are sensitive to rates
- The wealthy have lower mortality than population as a whole
- For past decade, SOI has used mortality rates for annuitants produced by Society of Actuaries
  - Annuitant 2000 mortality tables (A2000)

Mortality Rates

- Two new sources of mortality rates for the wealthy:
- Annuitant 2012 mortality tables (A2012)
  - Successor to the A2000 tables
  - Based on study from 2000-2004
- Saez-Zucman mortality rates (SZ)
  - Based on modelling wealth using capitalized income tax data
  - Linked to Social Security data to identify deaths

Mortality Rates



Mortality Rates

Period 2004-2008	Males			Females		
	Top 10%	Top 5%	Top 1%	Top 10%	Top 5%	Top 1%
Age 30-49	0.52	0.44	0.53	0.51	0.57	0.40
Age 50-64	0.61	0.53	0.43	0.67	0.57	0.71
Age 65-79	0.77	0.71	0.60	0.76	0.73	0.69
Age 80+	0.97	0.91	0.92	0.97	0.92	0.91
SZ Diff from NCHS	**0.58**	**0.55**	**0.56**	**0.56**	**0.53**	**0.53**

Source: Appendix C7, http://gabriel-zucman.eu/uswealth/

Preliminary Results

Figure 2a: Number of Top Wealth Holders with $2 Million or More in Assets, 2007

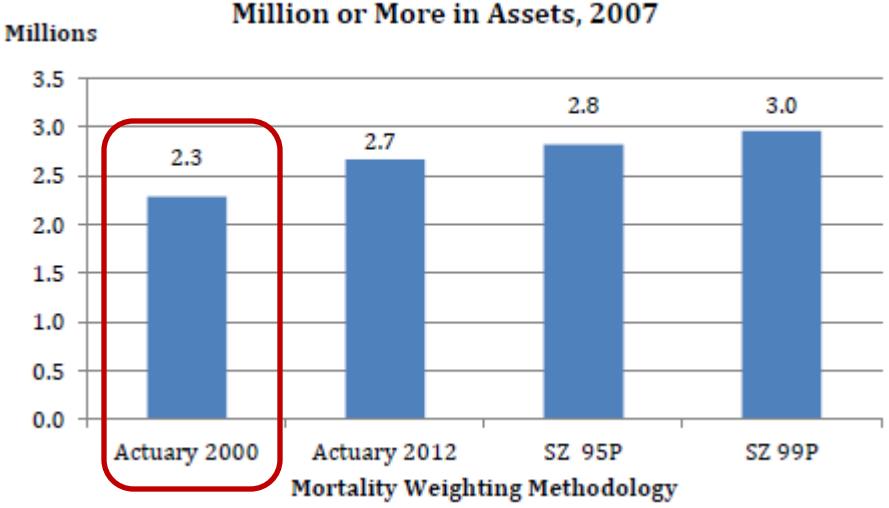
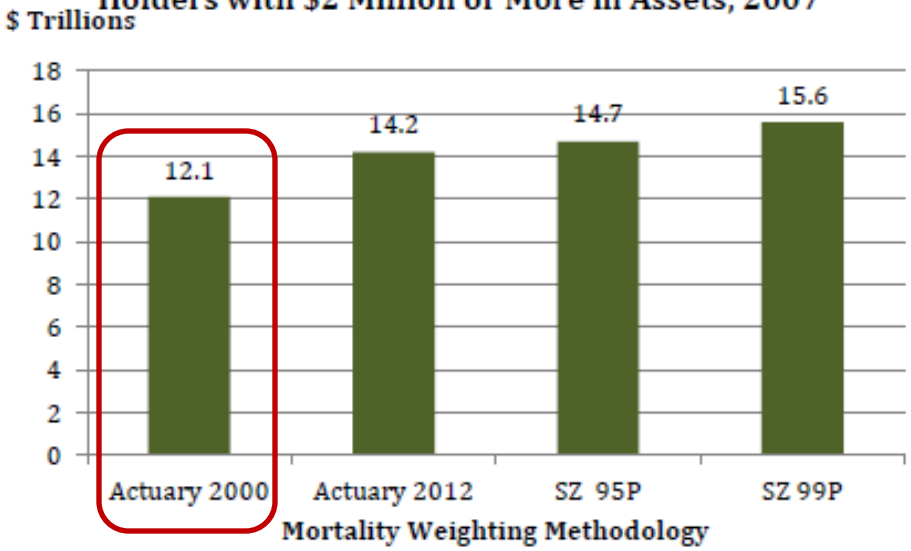
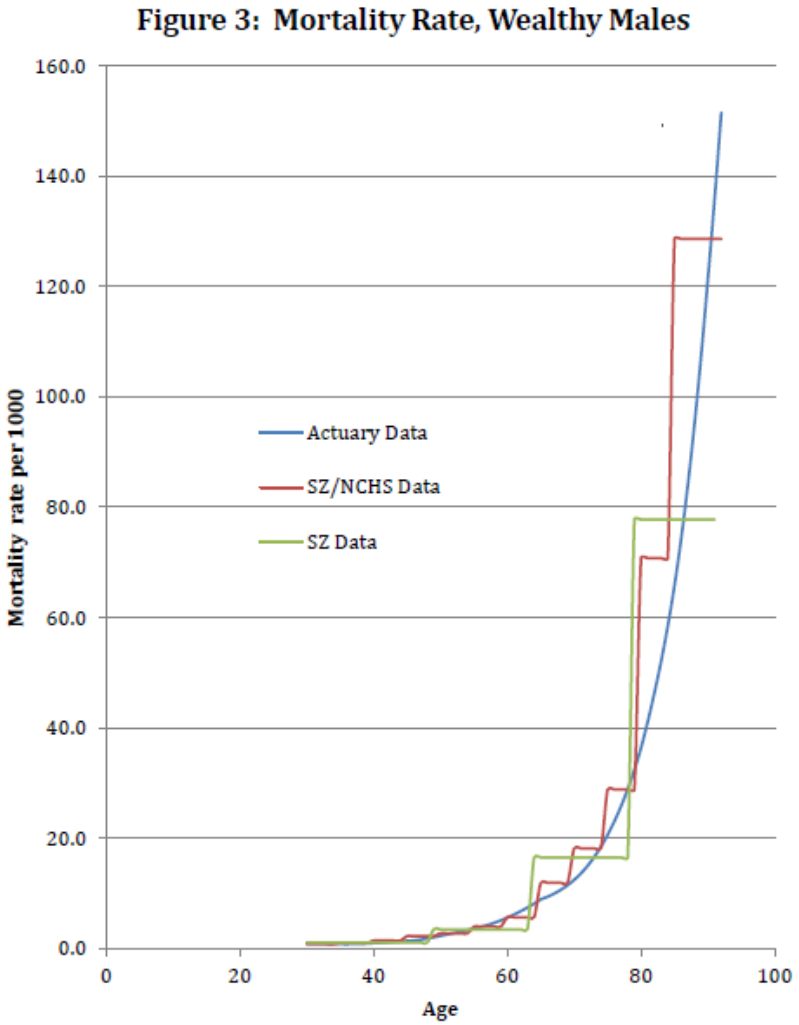


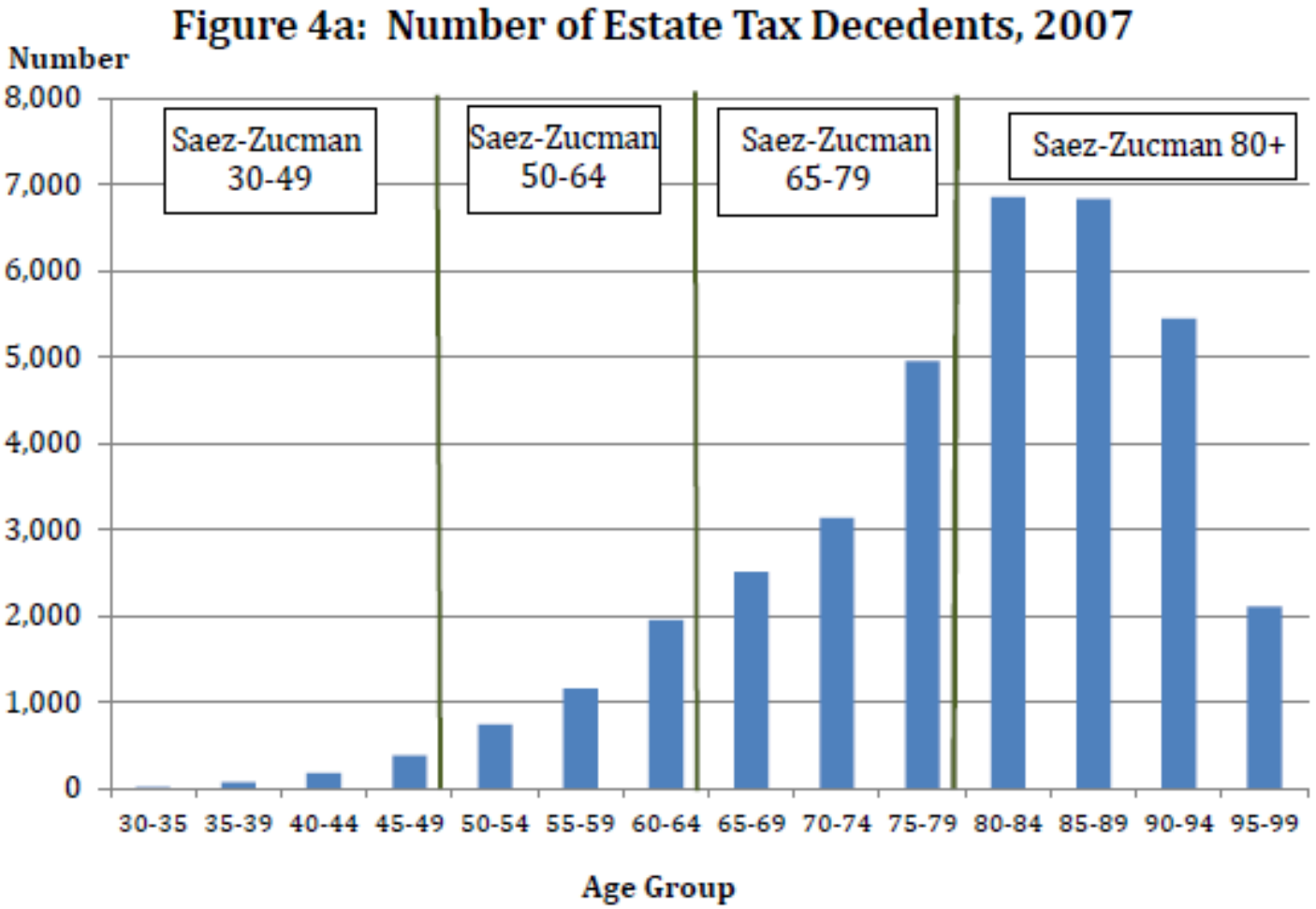
Figure 2b: Total Net Worth of Top Wealth Holders with $2 Million or More in Assets, 2007



Preliminary Results



Preliminary Results



Preliminary Results – Sex Distribution

Figure 5a: Age Distribution of Females, Actuary Data

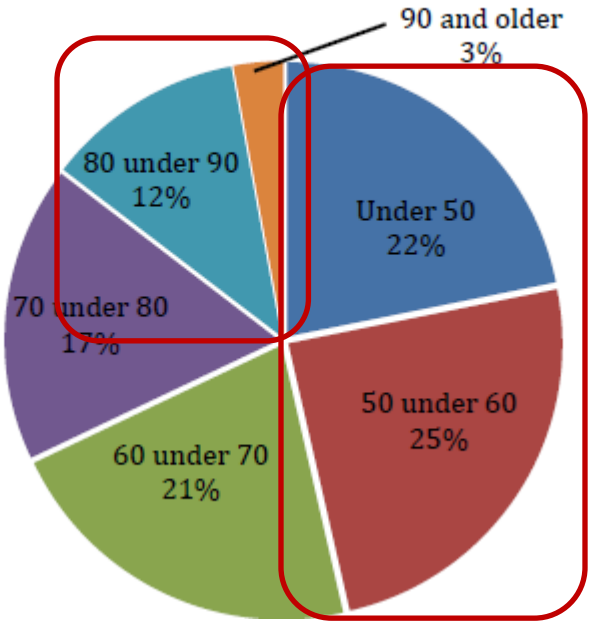
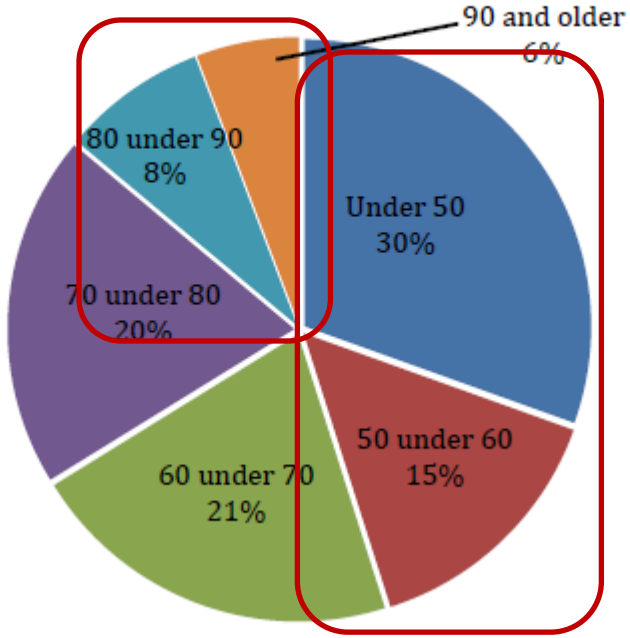
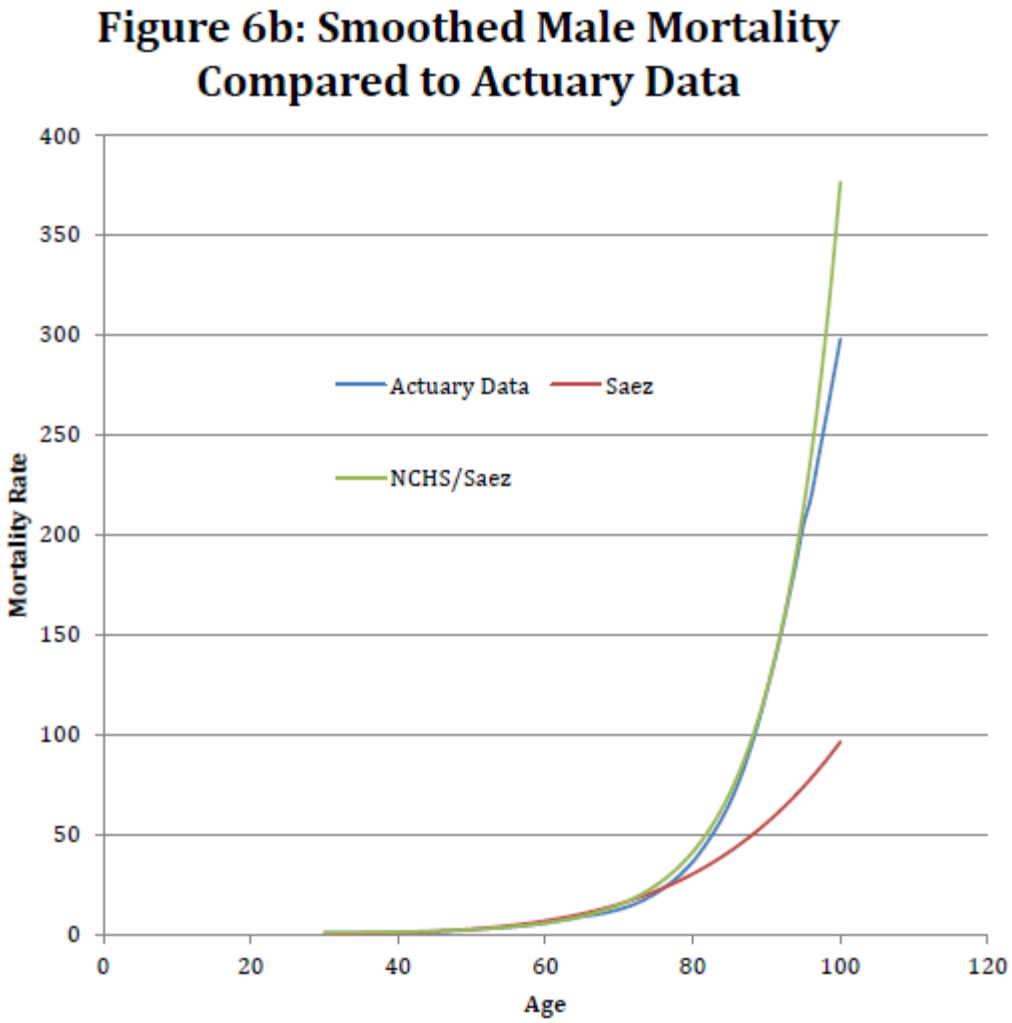


Figure 5b: Age Distrbution of Females, Saez-Zucman Data



Preliminary Results – Age Distribution for Males



Preliminary Results – Age Distribution for Females

Figure 7a: Number of Top Wealth Holders with $2 Million or More in Assets, 2007

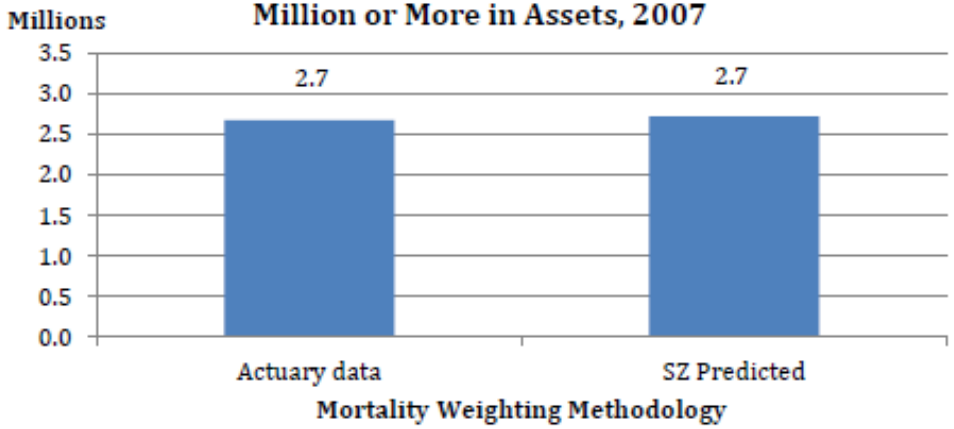
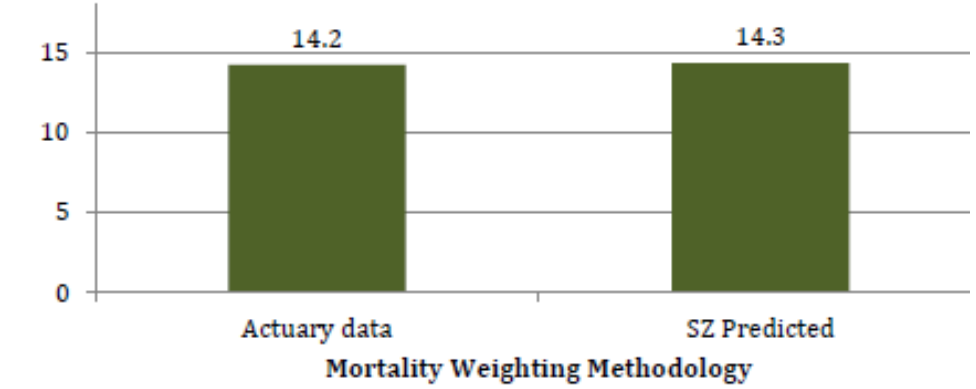


Figure 7b: Total Net Worth of Top Wealth Holders with $2 Million or More in Assets, 2007



Conclusion

- After smoothing to account for broad age categories, preliminary findings that Saez-Zucman and actuarial mortality rates produce very similar wealth estimates
- Use of broad age categories has significant impact on estimates
- Use of actuarial rates may be preferable due to less age bias

Panel Discussion

Are Piketty and Zucman Getting It Right?

Discussion Question

- Which approach to estimating mortality for the wealthy do you think is the most methodologically sound for SOI?

A PRODUCTIVE PARTNERSHIP, JOINT WORK WITH STANFORD

THE SOI-STANFORD COLLABORATION

*** DO NOT CITE OR QUOTE ***

PRESENTATION PREPARED FOR SOI PANEL MEETING, JUNE 5, 2015

DRAWING ON JOINT RESEARCH WITH PABLO MITNIK, MICHAEL WEBER, VICTORIA BRYANT, DAVID GRUSKY, MICHAEL HOUT, JONATHAN FISHER, MICHELLE JACKSON, DAVID JOHNSON, TIMOTHY SMEEDING, C. MATTHEW SNIPP, AMY O'HARA, J. TRENT ALEXANDER, & OTHERS

PARTIAL FUNDING FOR THIS RESEARCH CAME FROM THE U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (AE00101, OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION, AWARDED BY THE SUBSTANCE ABUSE MENTAL HEALTH SERVICE ADMINISTRATION), THE RUSSELL SAGE FOUNDATION, AND THE PEW CHARITABLE TRUSTS. THE OPINIONS EXPRESSED HERE ARE SOLELY THOSE OF THE AUTHORS AND DO NOT REPRESENT THOSE OF THE STANFORD CENTER ON POVERTY AND INEQUALITY, THE DEPARTMENT OF HEALTH AND HUMAN SERVICES, THE RUSSELL SAGE FOUNDATION, OR THE PEW CHARITABLE TRUSTS.

THREE SOI-STANFORD PROJECTS

NEW ESTIMATES OF INTERGENERATIONAL PERSISTENCE

EXPLOITING THE OCCUPATION FIELDS IN FORM 1040

BUILDING A NEW INTERGENERATIONAL PANEL

NEW ESTIMATES OF INTERGENERATIONAL PERSISTENCE

U.S. HAS LONG-STANDING INTEREST IN ECONOMIC MOBILITY … AND SOME TAX PROVISIONS ARE PARTLY JUSTIFIED AS MOBILITY-INCREASING (E.G., LOW-INCOME TAX CREDITS, ESTATE TAX)

IT IS ACCORDINGLY IMPORTANT TO MONITOR INTERGENERATIONAL ECONOMIC MOBILITY AND ASSESS WHETHER RAMPED-UP TAX POLICY MIGHT BE NEEDED

THE PROBLEM: THE ADMINISTRATIVE DATA REVOLUTION HAS NOT RESOLVED HOW MUCH INTERGENERATIONAL PERSISTENCE THERE IS

THE EVIDENCE DEFICIT

PROBLEM #1: A WIDE RANGE OF ESTIMATES (FROM SURVEY AND ADMINISTRATIVE DATA)

- SOLON'S (2008) META-ANALYSIS OF SURVEY AND ADMINISTRATIVE EVIDENCE: "INTER-GENERATIONAL EARNINGS ELASTICITY IN THE U.S. MAY WELL BE AS LARGE AS 0.5 OR 0.6"
- CHETTY ET AL. (2014): PREFERRED ESTIMATE OF 0.34 (FOR INTERGENERATIONAL INCOME ELASTICITY)
- A WIDE RANGE: ARE ONE-THIRD OF PERCENT INCOME DIFFERENCES TRANSMITTED FROM ONE GENERATION TO NEXT? OR IS IT TWO-THIRDS?

PROBLEM #2: PAUCITY OF TAX-RETURN EVIDENCE … BECAUSE OF METHODOLOGICAL PROBLEMS IN ESTIMATING INTERGENERATIONAL ELASTICITY (IGE)

WE NEED TO GET IT DONE: IGE IS KEY WORKHORSE MEASURE (EXPECTED PERCENT CHANGE IN CHILDREN'S INCOME GIVEN A ONE PERCENT INCREASE IN PARENTAL INCOME)

HOW DO WE GET IT DONE?

A SIMPLE GOAL: ESTABLISH HOW MUCH INTERGENERATIONAL PERSISTENCE THERE IS IN THE U.S. TODAY

TWO PROBLEMS NEED TO BE SOLVED

- DATA PROBLEM
- METHODS PROBLEM

ADDRESSING THE DATA PROBLEM

TABLE 1: CONSTRUCTION OF SOI-M PANEL

DATA	PURPOSE
TAX RETURNS FROM SOI FAMILY PANEL (1987-1996)	SOURCE OF PARENTAL INCOME DATA AND PARENT-CHILD SOCIAL SECURITY LINKS (WITH CLAIMED CHILDREN THEN TRACED FORWARD)
TAX RETURNS FROM THE REFRESHMENT SEGMENT OF THE OTA PANEL (1987-1996)	RECOVER "NONPERMANENT NONFILERS" (I.E., INDIVIDUALS IN 1987 NON-FILING POPULATION WHO APPEARED IN AT LEAST ONE 1988-96 RETURN)
POPULATION OF TAX RETURNS (1997-1998)	INCOME DATA FOR 1987 PARENTS (INCOME SECURED UP TO YEAR WHEN CHILD BECOMES 23 YEARS OLD)
POPULATION OF TAX RETURNS (1998-2010)	INCOME DATA FOR CHILDREN AND THEIR SPOUSES
W2 FORMS (1999-2010)	EARNINGS OF CHILDREN, INCLUDING NONFILING CHILDREN
1040SE FORMS (1999-2010)	SELF-EMPLOYMENT INCOME
SSA DATA MASTER FILE	DEMOGRAPHIC INFORMATION (AGE AND GENDER OF PARENTS AND CHILDREN, YEAR OF DEATH OF CHILDREN)
1099G FORMS	UNEMPLOYMENT INCOME OF NONFILING CHILDREN
CURRENT POPULATION SURVEY (CPS)	IMPUTED INCOME FOR NONFILING CHILDREN WITHOUT W-2 OR UI DATA (USING CPS FILING STATUS VARIABLES)

CHETTY ET AL.'S (2014) ANALYSIS OF 1996-2012 TAX DATA YIELDS CHILDREN 29-32 YEARS OLD IN 2011-12 … TOO EARLY IN CAREER TO YIELD GOOD IGE ESTIMATES?

SOLUTION: CONSTRUCT SOI-M PANEL

- POPULATION OF INTEREST: CHILDREN BORN 1972-75 WHO WERE LIVING IN U.S. IN 1987
- START WITH SOI 1987-96 FAMILY PANEL
- ADD OTA REFRESHMENT SEGMENT TO CAPTURE 1987 NONFILERS AND ADD 1997-98 IRTF DATA FROM CDW TO COMPLETE DATA FOR PARENTS
- USE CDW TO OBTAIN CHILDREN'S IRTF DATA FOR 1998-2010

WHY SOI-M PANEL IS SO ATTRACTIVE

- REDUCES LIFECYCLE BIAS BY EXAMINING MOBILITY OF CHILDREN AGES 35-38 IN 2010
- ADDRESSES ATTENUATION BIAS BY USING 9 YEARS OF PARENTAL INFORMATION

ADDRESSING THE METHODS PROBLEM

OLS LOG-LOG ESTIMATOR IS METHODOLOGICAL CONVENTION:

$$E(\ln Y \,|x) = \beta_0 + \beta_1 \ln x$$

TWO CHOICES – BOTH BAD – IF ONE OPTS FOR OLS LOG-LOG ESTIMATOR

- DROP CHILDREN WITHOUT EARNINGS OR INCOME → SELECTION BIAS (AN IGE THAT PERTAINS TO "WHEN THINGS ARE GOING WELL")
- KEEP CHILDREN WITHOUT EARNINGS OR INCOME AND ASSIGN ARBITRARY POSITIVE VALUE → ESTIMATES ARE EXTREMELY SENSITIVE TO CHOSEN VALUES

SOLUTION: DEFINE THE ESTIMAND CORRECTLY (IGE_e)

$$\ln E(Y|x) = \alpha_0 + \alpha_1 \ln x$$

ELIMINATES SELECTION BIAS AND EXTREME SENSITIVITY OF ESTIMATES

DATA AND METHODOLOGICAL FIXES REDUCE BIASES

LATE THIRTIES SAMPLE (VIA SOI-M PANEL) → REDUCES LIFECYCLE BIAS

NINE YEARS OF PARENTAL INFORMATION (VIA SOI-M PANEL) → REDUCES ATTENUATION BIAS

CORRECT ESTIMATOR → REDUCES SELECTION BIAS

RELAX CONSTANT-ELASTICITY ASSUMPTION → REDUCES FUNCTIONAL-FORM BIAS

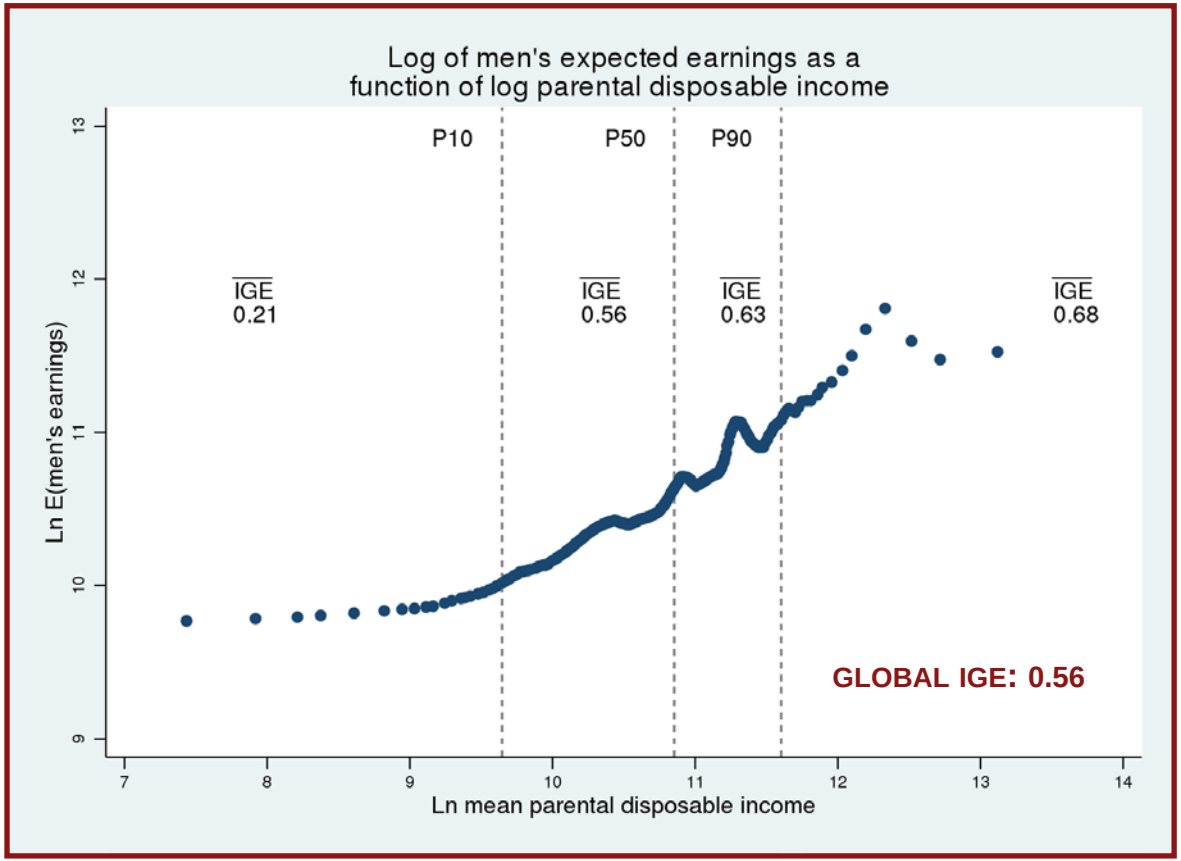
INCOME AND EARNINGS MEASURES

ANNUAL TOTAL INCOME OF PARENTS AND CHILDREN

ANNUAL AFTER-FEDERAL-TAX INCOME ("DISPOSABLE INCOME") OF PARENTS AND CHILDREN

INDIVIDUAL EARNINGS OF CHILDREN (INCLUDING EARNINGS FROM SELF-EMPLOYMENT)

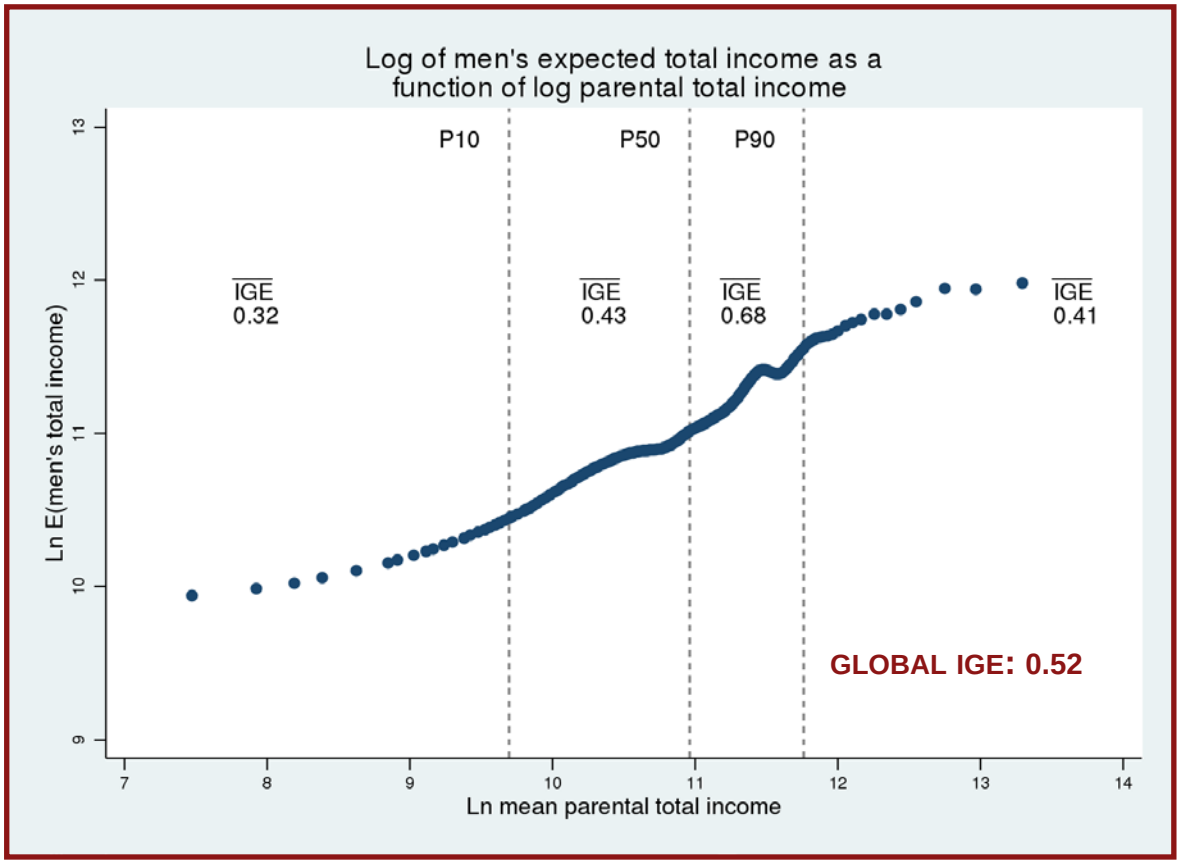
MEN'S EARNINGS CURVE



GLOBAL IGE, 0.56, IS AT UPPER END OF ESTIMATES

LEFT TAIL IS FLAT BUT THEN SLOPE INCREASES: CONVEX CURVE

MEN'S TOTAL INCOME CURVE



MEN'S TOTAL INCOME IGE, 0.52, AT UPPER END OF RANGE OF ESTIMATES

ANOTHER CONVEX CURVE WITH ELASTICITY BETWEEN P50 AND P90 SUBSTANTIALLY LARGER THAN THAT BETWEEN P10 AND P50

WOMEN'S TOTAL INCOME CURVE

GLOBAL IGE ONLY SLIGHTLY LOWER

AGAIN CONVEX: ELASTICITY BETWEEN P50 AND P90 SUBSTANTIALLY LARGER THAN BETWEEN P10 AND P50

CONCLUSION

IF ONE CORRECTS FOR SELECTION, ATTENUATION, LIFECYCLE, AND FUNCTIONAL-FORM BIASES

→ TOTAL INCOME IGEs AND MEN'S EARNINGS IGE ARE VERY HIGH … AND AT UPPER END OF EXISTING ESTIMATES

NEXT STEPS

BETTER DOCUMENTATION OF NEW SOI-M PANEL

UPDATE SOI-M PANEL WITH POST-2010 DATA

DEVELOP PROTOCOL FOR ANNUAL REPORTING OF KEY IGES BY SOI

THREE PROJECTS

NEW ESTIMATES OF INTERGENERATIONAL PERSISTENCE

EXPLOITING THE OCCUPATION FIELDS IN FORM 1040

BUILDING A NEW INTERGENERATIONAL PANEL

GOALS OF SOI-STANFORD OCCUPATION PROJECT

EXPLOIT OCCUPATION FIELDS ON FORM 1040 BY DEVELOPING CODING SCHEME AND APPLYING IT TO SOI SAMPLES

TAX POLICY USES: IMPROVED ESTIMATES OF INTERGENERATIONAL MOBILITY

- OCCUPATIONS CONVEY INFORMATION ON LIFETIME EARNINGS AND INCOME
- COMBINING ECONOMIC AND OCCUPATION REPORTS CORRECTS FOR UNDERESTIMATES OF INTERGENERATIONAL PERSISTENCE

CENSUS USES: OCCUPATION FIELDS FROM FORM 1040 MAY BE USEFUL FOR FILLING IN MISSING CENSUS, ACS, AND CPS REPORTS (ASSUMING REG CHANGES)

GENERAL SURVEY USES: ESTABLISH VIABILITY OF SHORT-RESPONSE OCCUPATION ITEMS

FIVE STEPS

STEP 1 (LINKING): LINK SELECTED FIELDS FROM 2011-12 TAX YEARS TO TRAINING SET DRAWN FROM 2011-12 CPS ASEC AND 2011-12 ACS (WHICH INCLUDE 2010 CENSUS OCCUPATION CODES)

STEP 2 (MACHINE LEARNING ON TRAINING SET): APPLY MACHINE LEARNING TO TRAINING SET TO DEVELOP CODING ALGORITHM USING VARIABLES ON FORM 1040, FORM W-2 , SSA MASTER FILE, AND OTHER SOURCES

STEP 3 (TEST AGAINST BALANCE OF DATA): TEST RESULTING PROTOCOL AGAINST BALANCE OF ASEC AND ACS DATA

STEP 4 (DEVELOP CODING SCHEME): IF RESULTS ARE SATISFACTORY, DEVELOP AGGREGATED VERSION OF 2010 OCCUPATION SCHEME THAT YIELDS ACCEPTABLY LOW ERROR RATES (ALSO ALLOW FOR MULTIPLE IMPUTATION)

STEP 5 (APPLY TO SOI AND POPULATION DATA): APPLY SCHEME TO CURRENT AND HISTORICAL SOI FILES AND RECENT POPULATION FILES

VARIABLES FOR MACHINE LEARNING

FORM 1040: NAME; ADDRESS; FILING STATUS; WAGES, SALARIES, & TIPS; BUSINESS INCOME; CAPITAL GAIN OR LOSS; RENTAL INCOME; FARM INCOME OR LOSS; UNEMPLOYMENT COMP.; SS BENEFITS; EDUCATOR EXPENSES; BUSINESS EXPENSES; STUDENT LOAN INTEREST DEDUCTION; TUITION AND FEES; PAID PREPARER FLAG

SCHEDULES A, C, D, E: UNREIMBURSED EMPLOYEE EXPENSES; NAME OF PROPRIETOR; PRINCIPAL BUSINESS; BUSINESS NAME; EIN; GROSS RECEIPTS; EXPENSES FOR BUSINESS USE OF HOME; NET SHORT TERM GAIN OR LOSS FROM PARTNERSHPS, S CORPORATIONS, ESTATES AND TRUSTS; RENTS AND ROYALTIES

FORM 1099-MISC: PAYER'S NAME, ADDRESS, FEDERAL ID NUMBER; RENTS; ROYALTIES

FORM W-2: EMPLOYER NAME AND ADDRESS, EIN; INDUSTRY

SSA MASTER FILE: GENDER; AGE

VARIABLES FROM PRIOR FILING YEARS: LAST YEAR'S OCCUPATION (AND MANY OTHERS)

THREE PROJECTS

NEW ESTIMATES OF INTERGENERATIONAL PERSISTENCE

EXPLOITING THE OCCUPATION FIELDS IN FORM 1040

BUILDING A NEW INTERGENERATIONAL PANEL

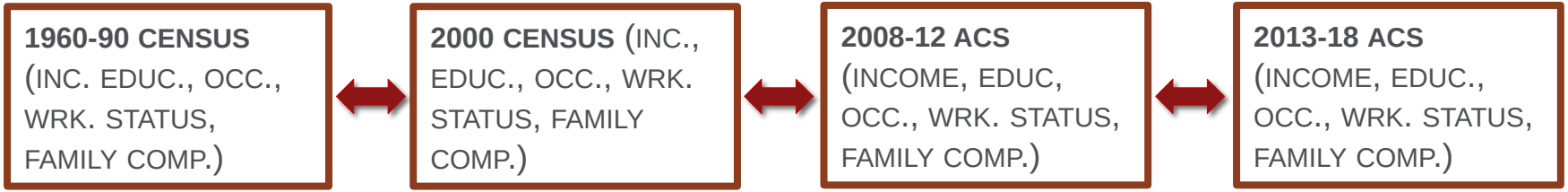
THE AMERICAN OPPORTUNITY STUDY (AOS)

THE U.S. HAS AN *UNASSEMBLED PANEL …* AND THE AMERICAN OPPORTUNITY STUDY (AOS) IS A NEW INITIATIVE TO ASSEMBLE IT

ALTHOUGH TAX DATA ARE KEY RESOURCES IN ADDRESSING LABOR MARKET ISSUES, THE AOS WOULD ALLOW US TO BETTER ADDRESS PROBLEMS ARISING FROM NONFILING AND MISSING DATA (E.G., RACE)

AND OF COURSE TAX DATA CAN ONLY BE USED FOR ANALYSES DIRECTLY RELEVANT TO TAX POLICY AND TAX ADMINISTRATION

STEP #1: LINKING RECORDS FROM ACS AND CENSUS ACROSS YEARS

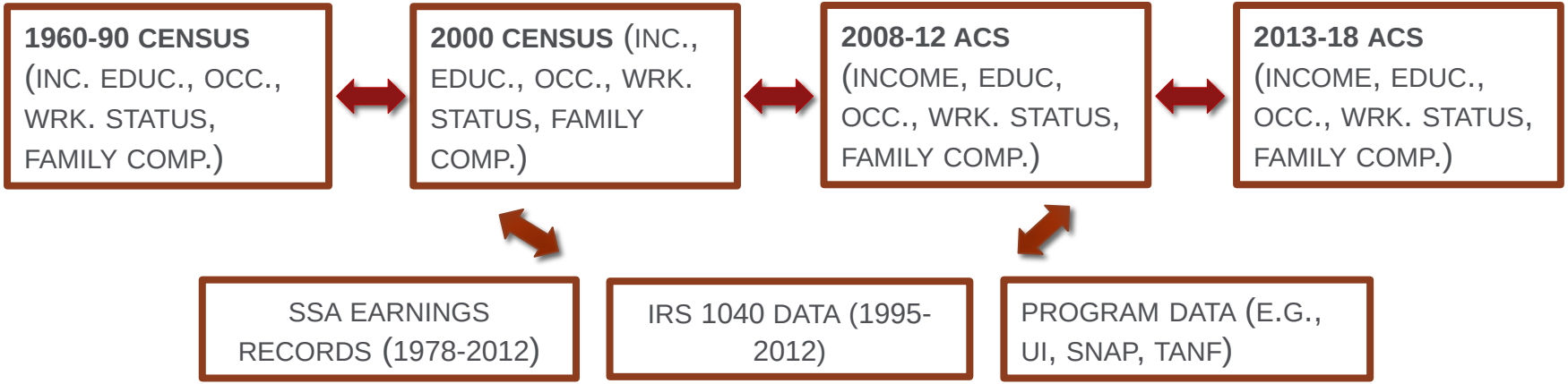


ASSIGN PROTECTED IDENTIFICATION KEYS (PIKS) TO RECORDS

USE VARIABLES FROM ACS OR CENSUS LONG FORM (E.G., FIRST NAME, LAST NAME, YEAR OF BIRTH, ADDRESS, SEX) TO FIND SSN IN SSA NUMIDENT FILE

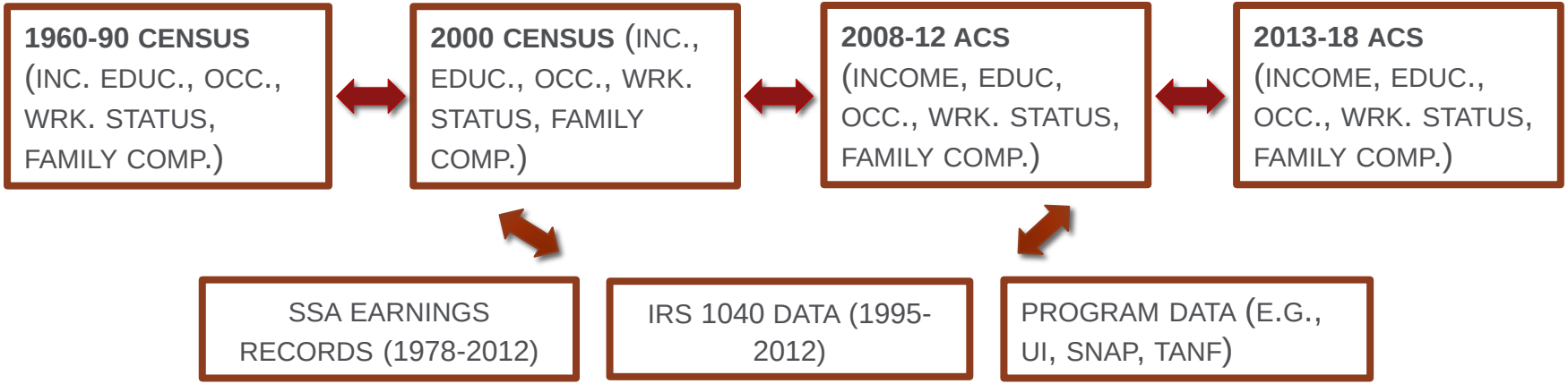
RESULT: INSTANT PANEL DATA THAT IS AUTOMATICALLY REFRESHING (I.E., NO TOP-DOWN, ARTIFICIAL DEFINITION OF POPULATION AT POINT OF CONSTRUCTION)

STEP #2: ADDING IN ADMINISTRATIVE DATA



ONCE PIKS ARE ASSIGNED, LINKAGES TO ADMINISTRATIVE DATA CAN ALSO BE MADE (CONDITIONAL OF COURSE ON APPROVALS TO DO SO)

STEP #3: LINKING CHILDREN WITH PARENTS



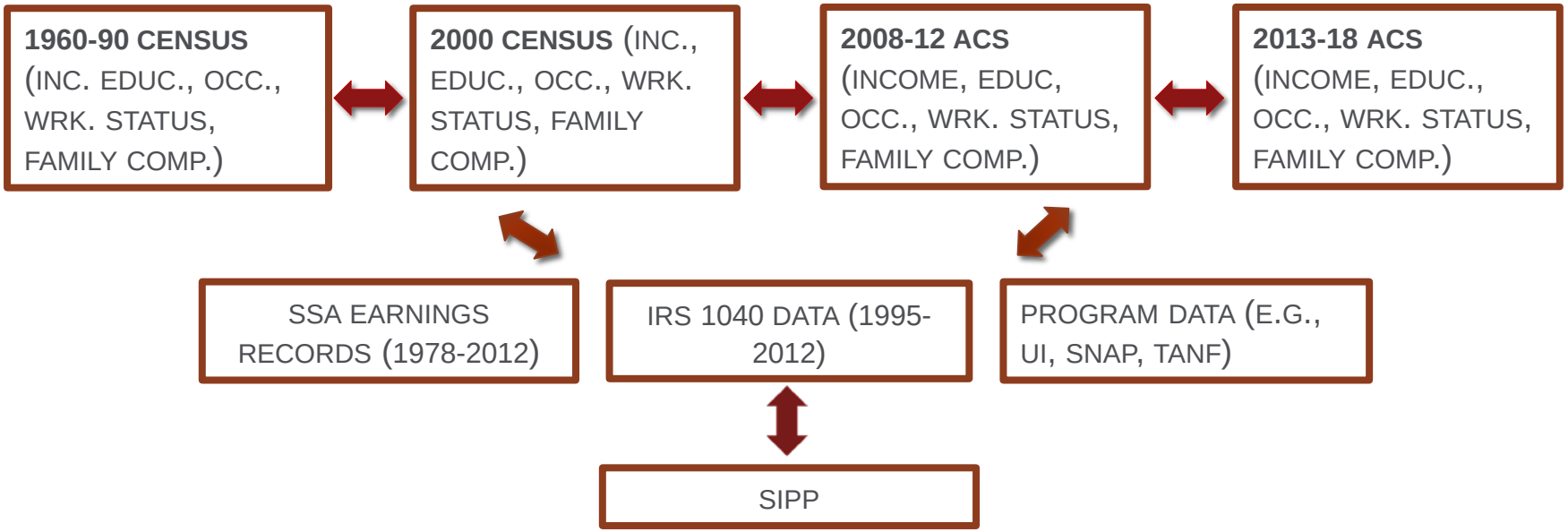
KIDLINK FILES (E.G., FORM SS-5 IF HISTORICAL FORMS AVAILABLE)

CORESIDENCY IN ACS, CENSUS LONG FORM

PARENTAL REPORTS OF CHILDREN'S SSN TO IRS

THE RESULT: AOS

STEP #4: SLIPPING IN THE SURVEY



SURVEYS WITH IDENTIFIERS CAN BE SLIPPED IN

THE SURVEY AS A LEAN AND MEAN VALUE-ADDED INSTRUMENT

CRITICISMS AND CHALLENGES

"THE AOS IS LARGE AND UNGAINLY AND WILL SINK UNDER ITS OWN WEIGHT"

"THE AOS IS A REGISTER ON THE SLY"

"THE FORMIDABLE SECURITY ISSUES WILL BE THE UNDOING OF THE AOS"

"THE AOS POPULATION IS POORLY SPECIFIED"

"APPROVALS TO LINK TO ADMINISTRATIVE DATA WILL NEVER BE SECURED"

“THE AOS IS A REGISTER ON THE SLY”

TECHNICAL RESPONSE: ON-DEMAND DATA LINKAGE SERVICE … NOT A REGISTER

HEAD-ON RESPONSE: THE PAPERWORK REDUCTION ACTS OF 1980 AND 1985 *MANDATE* THE AOS

"FORMIDABLE SECURITY ISSUES WILL BE UNDOING OF AOS"

FIRST-ORDER CONCERNS: LEGITIMATE THREATS TO SECURITY

SECOND-ORDER CONCERNS: FALLOUT FROM UNWARRANTED PUBLIC WORRIES

FIRST-ORDER CONCERNS

IDENTIFIERS ARE JUST PRODUCTION TOOLS … NO NEW CONCERNS

RELEASED TO CAREFULLY VETTED RESEARCH AND RESEARCHERS … NO NEW CONCERNS

ANALYZED IN RDCs OR, IN THE CASE OF ESPECIALLY SENSITIVE DATA, RDCs ON STEROIDS … NO NEW CONCERNS

CANNOT RULE OUT LEGITIMATE WORRIES (AND HENCE OPEN DISCUSSION IS NEEDED)

SECOND-ORDER CONCERNS

STANDARD PRESCRIPTION FOR MISINFORMATION: OPEN DISCUSSION

NEXT STEPS

DIGITIZING 1960-90 CENSUSES

IMPROVING PIKING METHODOLOGY

IMPROVING INTERGENERATIONAL LINKAGES

REDUCING SECURITY CONCERNS

FUNDING

PAYOFF TO AOS

IMPROVED EVIDENCE ON LABOR MARKET OUTCOMES AND INTERGENERATIONAL MOBILITY

LOW COST POLICY AND PROGRAM EVALUATION

REDUCED RELIANCE ON SURVEYS

CROSS-SOURCE MISSING DATA FILL-INS

SAMPLING FRAME

Panel Discussion

A Productive Partnership, Joint Work With Stanford

AN OVERVIEW OF THE SOI CONSULTANTS PANEL

Panel Discussion

An Overview of the SOI Consultants Panel

The SOI Panel

What does the SOI Panel look like?

Background
Expertise
Service

Survey responses of 13 panel members

What does the SOI Panel look like?

Background
Expertise
Service

2. Current position (required)

#	Answer		Response	%
1	Academic		4	31%
2	US federal government		0	0%
3	Other government		1	8%
4	Non-profit		5	38%
5	Private industry		2	15%
6	Retired		1	8%
	Total		13	100%

15 invitations to the survey

4. Have you ever worked in the US federal government? (required)

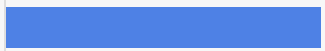
#	Answer		Response	%
1	Yes		12	92%
2	No		1	8%
	Total		13	100%

5. If yes, please specify (check all that apply):

#	Answer		Response	%
1	Executive Branch: Treasury (OTA)		8	67%
2	Executive Branch: IRS		2	17%
3	Executive Branch: Commerce (including Census)		2	17%
4	Executive Branch: Other		2	17%
5	Legislative Branch: JCT		4	33%
6	Legislative Branch: CBO		1	8%
7	Legislative Branch: Other		0	0%
8	Federal Reserve		1	8%

Treasury “other”, and other agencies (e.g., NSF) omitted

9. Research area(s) / areas of expertise: (required - check all that apply)

#	Answer		Response	%
1	Individual		10	77%
2	Corporate		6	46%
3	Partnership		3	23%
4	International		4	31%
5	Estate & Gift		6	46%
6	State & Local		4	31%
7	Tax-exempt		3	23%
8	Other		3	23%

Statistic	Value

Tax Administration
Statistics
record linkage and Mathematical Statistics

11. How many years have you been a member of the panel? (required)

#	Answer		Response	%
1	1-3		0	0%
2	4-6		4	31%
3	7-9		1	8%
4	10-12		1	8%
5	13-15		1	8%
6	15-20		0	0%
7	More than 20		6	46%
	Total		13	100%

More than 185 total years of Panel participation
Average tenure greater than 14 years

13. Prior to joining the panel, did you have access to, and experience with, non-public SOI data? (required)

#	Answer		Response	%
1	Yes		11	85%
2	No		2	15%
	Total		13	100%

14. Since joining the panel, have you had access to, and experience with, non-public SOI data? (required)

#	Answer		Response	%
1	Yes		10	77%
2	No		3	23%
	Total		13	100%

15. Have you ever published an article in the SOI Bulletin? (required)

#	Answer		Response	%
1	Yes		3	23%
2	No		10	77%
	Total		13	100%

17. Have you ever co-authored a paper with a SOI staff member? (required)

#	Answer		Response	%
1	Yes		7	54%
2	No		6	46%
	Total		13	100%

. What do you think is the optimal size of the panel (please enter a whole number)?

. What do you think is the optimal size of the panel (please enter a whole number)?

Text Response
Under 20, larger than 7 or 8.
15+
12
8
12-15
Aproximately 10
A prime number 11 to 17, with members appointed for a fixed number of years
15
10 (or so).
10
12

Estimated mean = 12-ish

16. Do you think panel members should be appointed to serve a fixed-length term? (required)

#	Answer		Response	%
1	Yes		6	46%
2	No		7	54%
	Total		13	100%

27. If terms were fixed, how many years should a term be? (please enter a whole number)

Text Response

27. If terms were fixed, how many years should a term be? (please enter a whole number)

Text Response
Not sure about fixed terms. There is the prior matter of overlapping terms of appointees that is not dealt with in the questionnaire.
5 years
5 to six years
3/5 years
6
Maybe 5 (with rotating terms).
10 years
4
5

Estimated mean = 6-ish

20. If panel members were appointed to fixed terms, should the appointment be renewable?

#	Answer		Response	%
1	Yes		10	83%
2	No		2	17%
	Total		12	100%

26. Do you think that the panel should have a non-IRS chair or co-chair?

#	Answer		Response	%
1	Yes		12	100%
2	No		0	0%
	Total		12	100%

Open-ended Questions

Why do you serve on the panel

What characteristics/qualifications are necessary

What responsibilities should members have

Minimum commitments

Process to identify new members

Metrics to judge the panel's effectiveness

Challenges in being effective

Other thoughts

IRS

2015 Consultants Panel Meeting

Closing Remarks

Research, Analysis & Statistics

STATISTICS OF INCOME