On the Nature of Entrepreneurship

A. Bhandari,[1] T. Kass,[2] T. May,[3] E. McGrattan,[1] and E. Schulz[4]

[1]Department of Economics
University of Minnesota

[2]Office of Tax Analysis
Department of Treasury

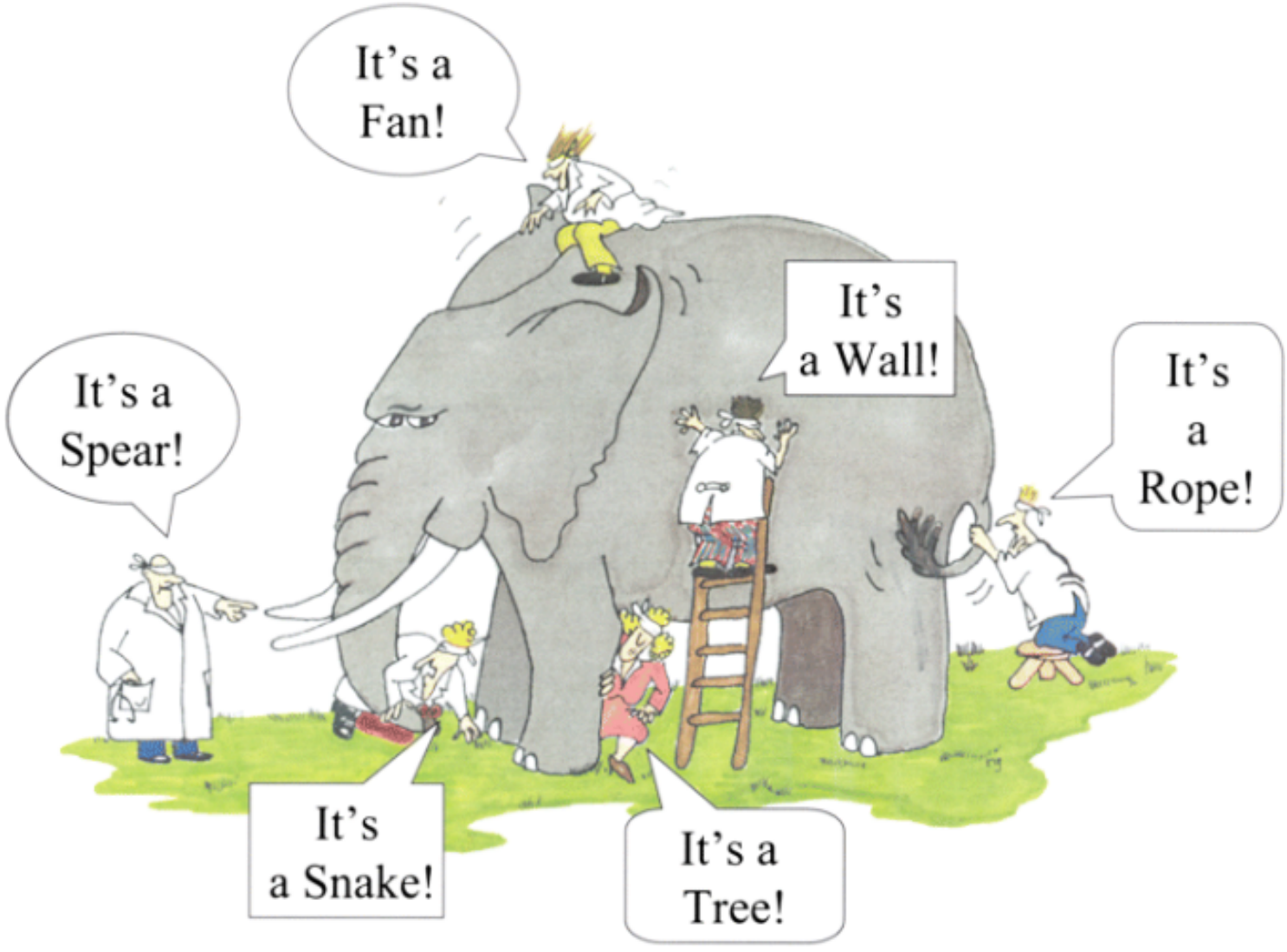
[3]Department of Economics
California State University, Fullerton

[4]Internal Revenue Service
Department of Treasury

Disclaimer

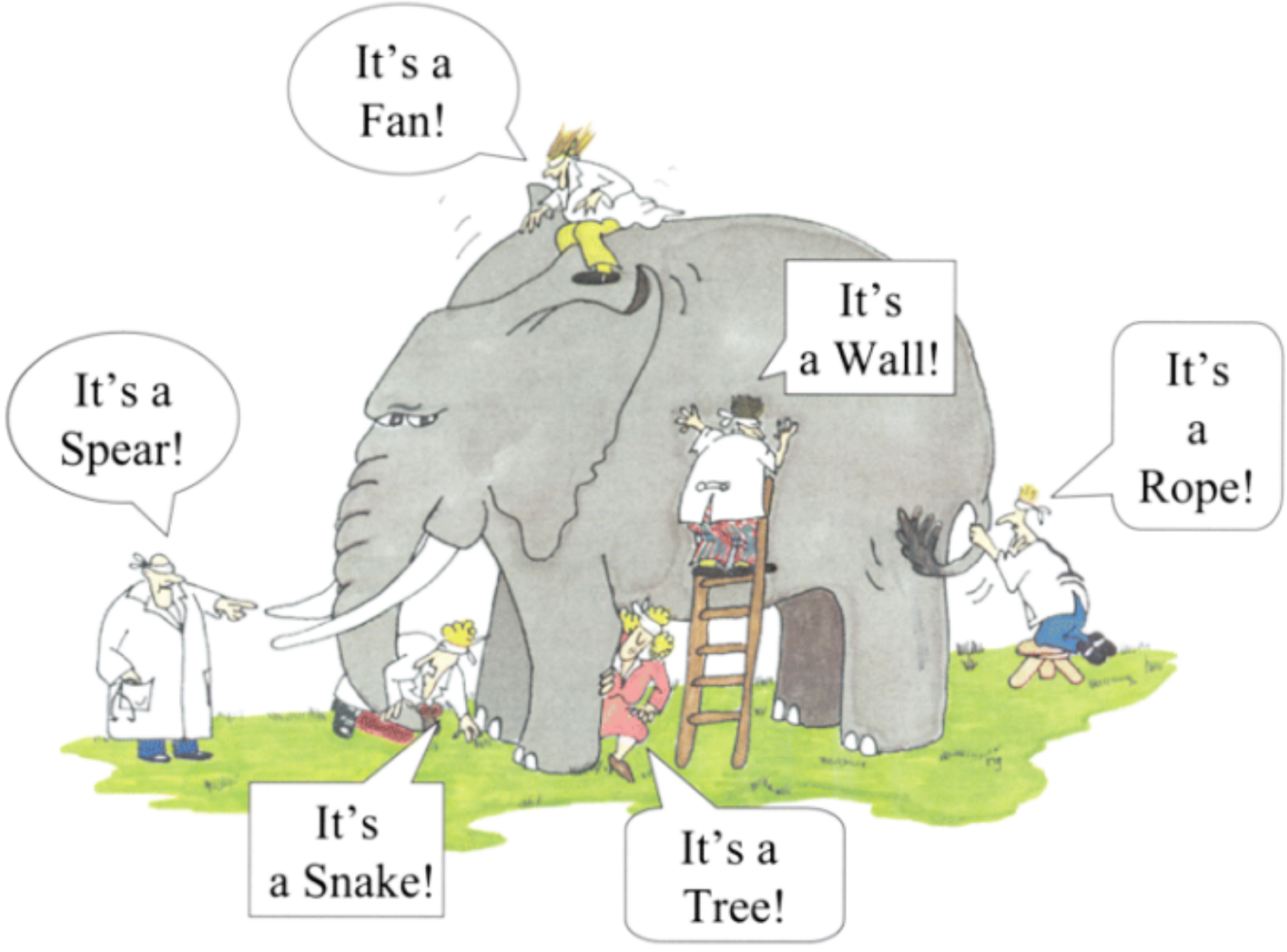
All data work for this project involving confidential taxpayer information was done at IRS facilities, on IRS computers, by IRS and Department of the Treasury employees, and at no time was confidential taxpayer data ever outside of the IRS computing environment. The findings, interpretations, and conclusions expressed in the paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury or the Internal Revenue Service. All results have been reviewed to ensure that no confidential information is disclosed.

Six Blind Men and an Elephant



~~Six Blind Men~~ and ~~an Elephant~~

(surveys) (entrepreneurship)



This Paper

- Provides new insights on the nature of entrepreneurship
- How?
  - Assemble novel longitudinal database of business owners
  - Study patterns of life-cycle income profiles
  - Analyze determinants of entrepreneurial choice

New Insights

- Contrast IRS and survey data
- Reevaluate central hypotheses regarding:
  - Non-pecuniary motives
  - Idiosyncratic and aggregate risk
  - Business dynamism trends
  - Liquidity constraints

Data

Primary Data Source

- Administrative IRS individual and business tax filings
- Samples of living individual with US SSN, 2000-15
  - All individuals aged 25–65
  - Balanced panel born 1950–1975
- Income measures:
  - Self-employment (SE) income
    - Schedule C net profits
    - Schedule K-1 ordinary business income
    - W-2 wages of S corporation owners
  - Paid-employment (PE) income
    - W-2 wages of non-owners

Analysis of Two Samples

- Contrast IRS versus CPS with simple moments for ages 25-65
  - Assign employment status each year with just incomes:
    - Self-employed (SE) if |SE income| >\$5,000 and > PE income
    - Paid-employed (PE) if not SE, PE income > \$5,000
    - Not-employed (NE) if not SE, PE

  $\Rightarrow$ Main motivation for longitudinal study

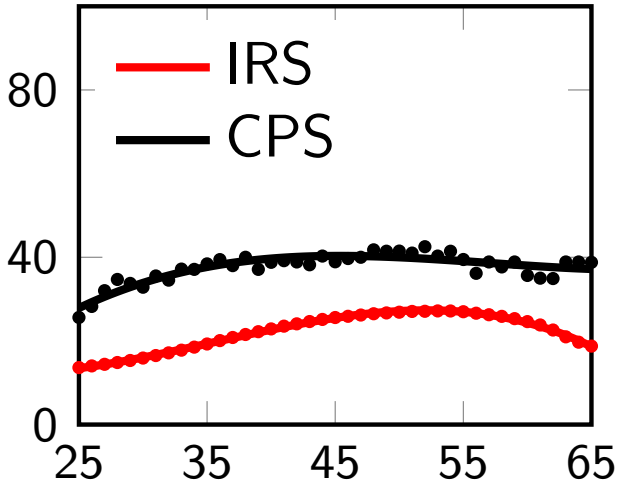
- Reevaluate hypotheses with longitudinal IRS balanced panel
  - Relax SE criteria and include individuals with
    - Share of gross profits >PE income
    - Share in business $\times$ employees ≥ 1

  $\Rightarrow$ Want to include SE running businesses but making little

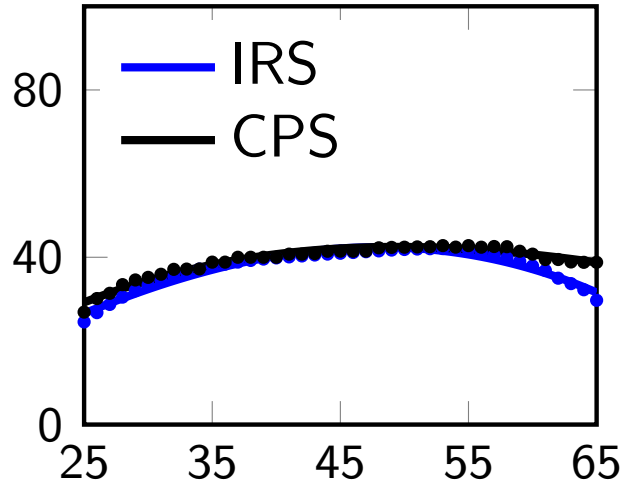
Comparison of Simple Moments: IRS versus CPS

IRS v. CPS (Thous. 2012$)

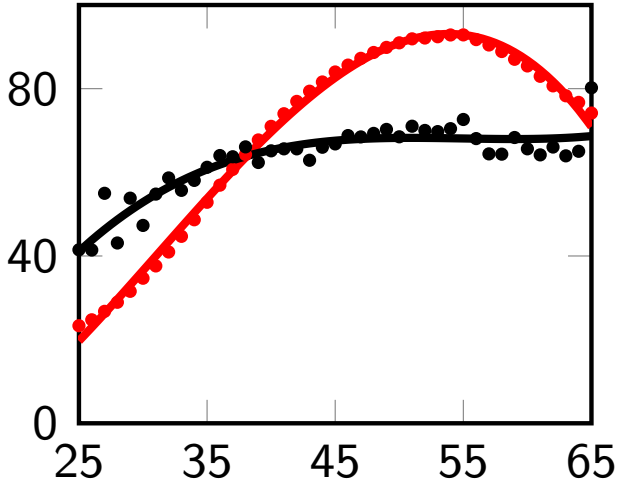
SE Median Income



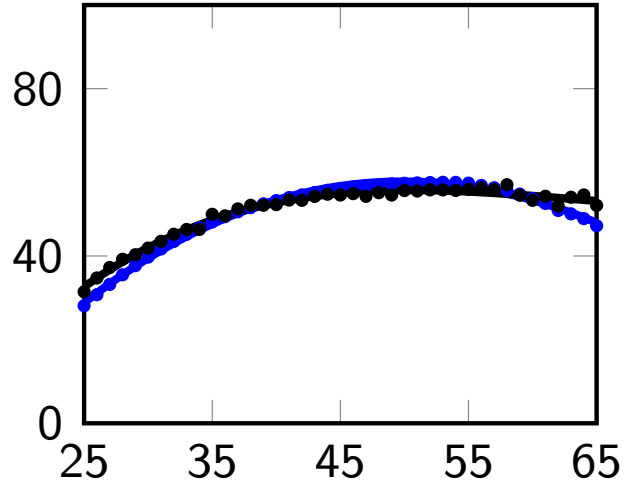
PE Median Income



SE Mean Income



PE Mean Income



IRS v. CPS Takeaways

- IRS and CPS close for paid-employed
- CPS misses self-employed, especially
  - High mid-career earners in key industries (prof services, health)
  - Low early-career earners with business losses

Income Cutoff	CPS Share	IRS Share
$< \$8{,}500$	0.2	-8.1
$> \$100{,}000$	53.0	80.2

- Comparison even worse if misreporting adjustments included

Reevaluate Central Hypotheses

Role for Non-pecuniary Motives

- Hypothesis: SE earn less than PE peers but compensated for
  - Being their own bosses
  - Having flexible schedules
- Survey evidence: SE discount is $\approx 35\%$
- IRS evidence:
  - Estimate life-cycle income profiles from longitudinal panel
  - Use detailed information from tax filings to group peers

Estimated Life-Cycle Income Profiles

$$y_{it} = \underbrace{\alpha_i}_{\substack{\text{fixed} \\ \text{effects}}} + \underbrace{\beta_{g(i),t}}_{\substack{\text{time} \\ \text{effects}}} + \underbrace{\sum_{a=a_0}^{a(i,t)} \gamma^a_{c(i),g(i)}}_{\substack{\text{age} \\ \text{effects}}} + \epsilon_{it}$$

where

- $i \in \mathcal{I}$ is set of individuals
- $t \in \mathcal{T}$ is set of calendar years
- $c \in \mathcal{C}$ is set of birth years
- $a \in \mathcal{A}$ is set of ages
- $g \in \mathcal{G}$ is set of groups partitioning $\mathcal{I}$

Identification

- Estimation of time ($\Delta\beta$) and age (γ) effects:

$$\Delta y_{it} = \underbrace{\Delta\beta_{g(i),t} + \gamma^{a(i,t)}_{c(i),g(i)}}_{\text{identification}} + \Delta\epsilon_{it}$$

- Identification:
  - Assume age effects are constant across binned cohorts
  - Normalize time effects to reflect group-specific growth

Groups

- Time-invariant characteristics include usual ones:
  - Cohort and gender (from SSA link to IRS)
  - Primary industry, marital status, children (on tax files)
  - Imputations of skills/education (from occupation on tax files)
- Plus partitions of sample based on employment history
  - Split full sample into primarily-employed or not
  - Split primarily-employed into
    - *Primarily paid-employed*
    - *Tried self-employment*

$\Rightarrow$ set $\mathcal{G}$ has 32,256 subgroups

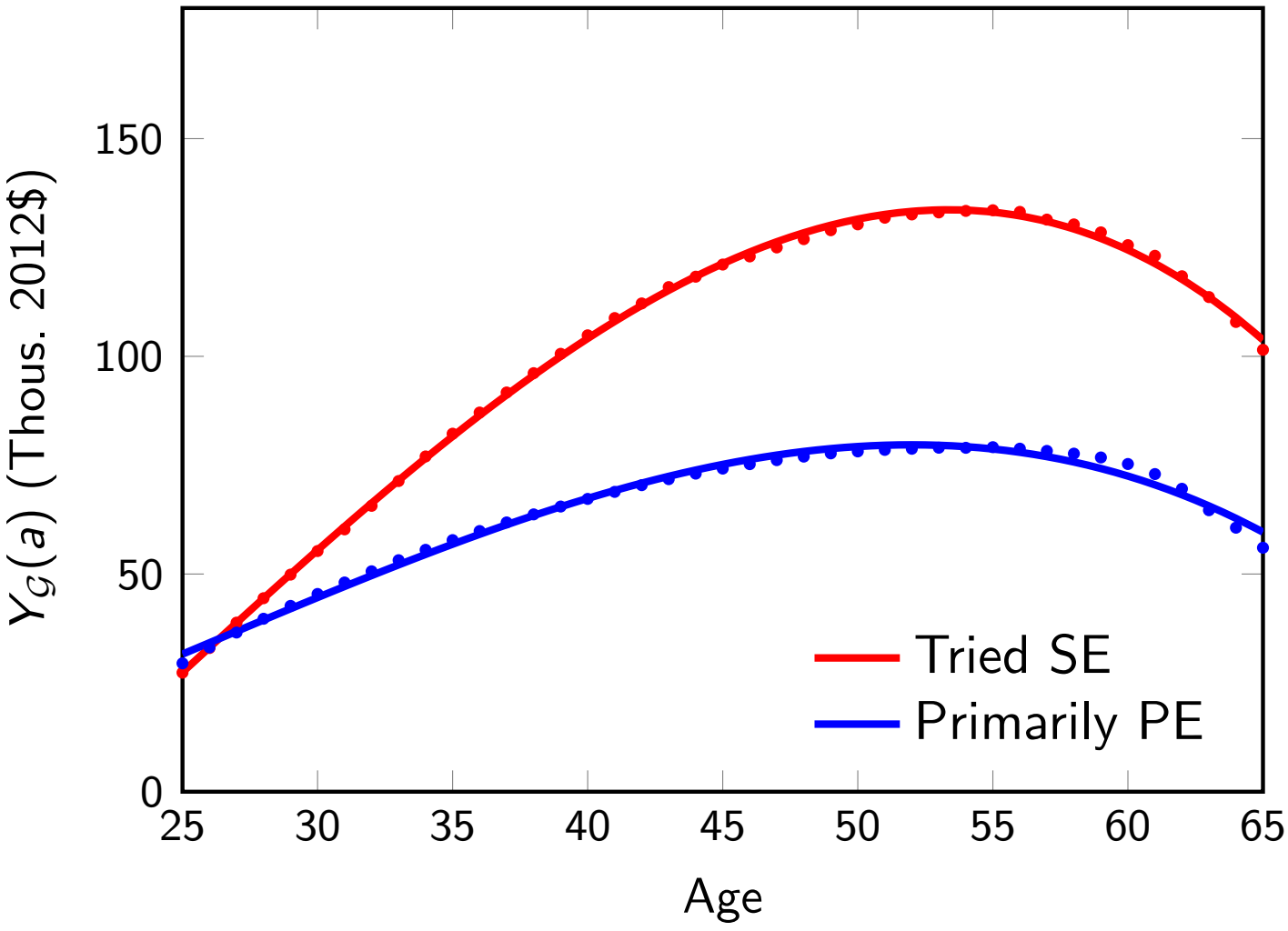
Summarizing Key Parameter Estimates

- For group g and age a:
  - N_g^a= number of individuals
  - $\bar{\gamma}_g^a$= age effect for g modified to include cohort bin
  - $\overline{\Delta\beta_g}$= average time effects
- Define integrated income at age a as:

$$Y_{\mathcal{G}}(a) = \underbrace{Y_{\mathcal{G}}(25)}_{\substack{\text{average} \\ \text{at 25}}} + \underbrace{\sum_{j=26}^{a} \sum_{g \in \mathcal{G}} \frac{N_g^j}{\sum_k N_k^j} (\bar{\gamma}_g^j + \overline{\Delta\beta_g})}_{\text{estimated age plus time effects}}$$

$\Rightarrow$ Next, plot $Y_{\mathcal{G}}$, $\mathcal{G}$=SE,PE (see appendix for subgroups)

Estimated Income Profiles



Main Takeaway from IRS Income Profiles

- Tried SE group has
  - Higher average income
  - Steeper growth profiles

  $\Rightarrow$ No evidence of SE discount

- Next, look for evidence in relevant subsamples

Investigating the Hypothesis

- Construct sample that fits Hamilton-Hurst-Pugsley narrative
  - Split *primarily SE* into those earning above/below PE peers
  - Count heads and incomes for each subgroup
  - Repeat exercise after imputing underreported incomes

Investigating the Hypothesis

- Construct sample that fits Hamilton-Hurst-Pugsley narrative
  - Split *primarily SE* into those earning above/below PE peers
  - Count heads and incomes for each subgroup
  - Repeat exercise after imputing underreported incomes
- Main findings w/o and w/ imputations:
  - w/o: 57% earn less than PE and 16% of income
  - w/: 37% earn less than PE and 10% of income

Investigating the Hypothesis

- Construct sample that fits Hamilton-Hurst-Pugsley narrative
  - Split *primarily SE* into those earning above/below PE peers
  - Count heads and incomes for each subgroup
  - Repeat exercise after imputing underreported incomes
- Main findings w/o and w/ imputations:
  - w/o: 57% earn less than PE and 16% of income
  - w/: 37% earn less than PE and 10% of income
- Punchline:
  - Surveys have typical individual in SE, not typical $ in SE

Role for Idiosyncratic Risk

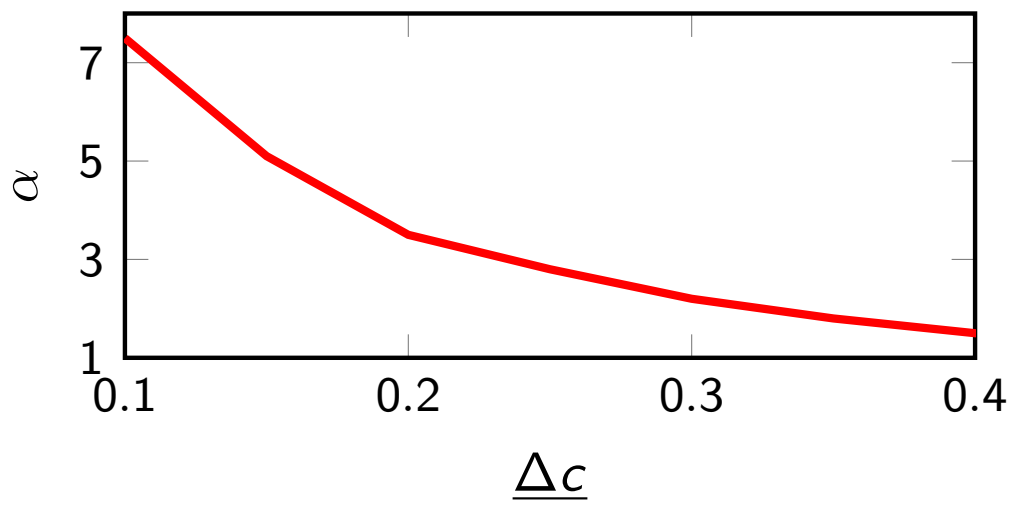
- Puzzle: SE returns too low to compensate for risk
- Survey evidence: comparisons of private returns to stocks
- IRS evidence: data
  - Variability and persistence of income growth for SE,PE
  - External sources of insurance for consumption risk

Investigating the Puzzle

- Use IRS data to parameterize consumption risk model with:
  - Epstein-Zin preferences and risk aversion α
  - Insurance on lower bound of consumption growth $\underline{\Delta c}$
- Trace out SE-PE indifference curves varying α, $\underline{\Delta c}$

Investigating the Puzzle

- Use IRS data to parameterize consumption risk model with:
  - Epstein-Zin preferences and risk aversion α
  - Insurance on lower bound of consumption growth $\underline{\Delta c}$
- Trace out SE-PE indifference curves varying α, $\underline{\Delta c}$

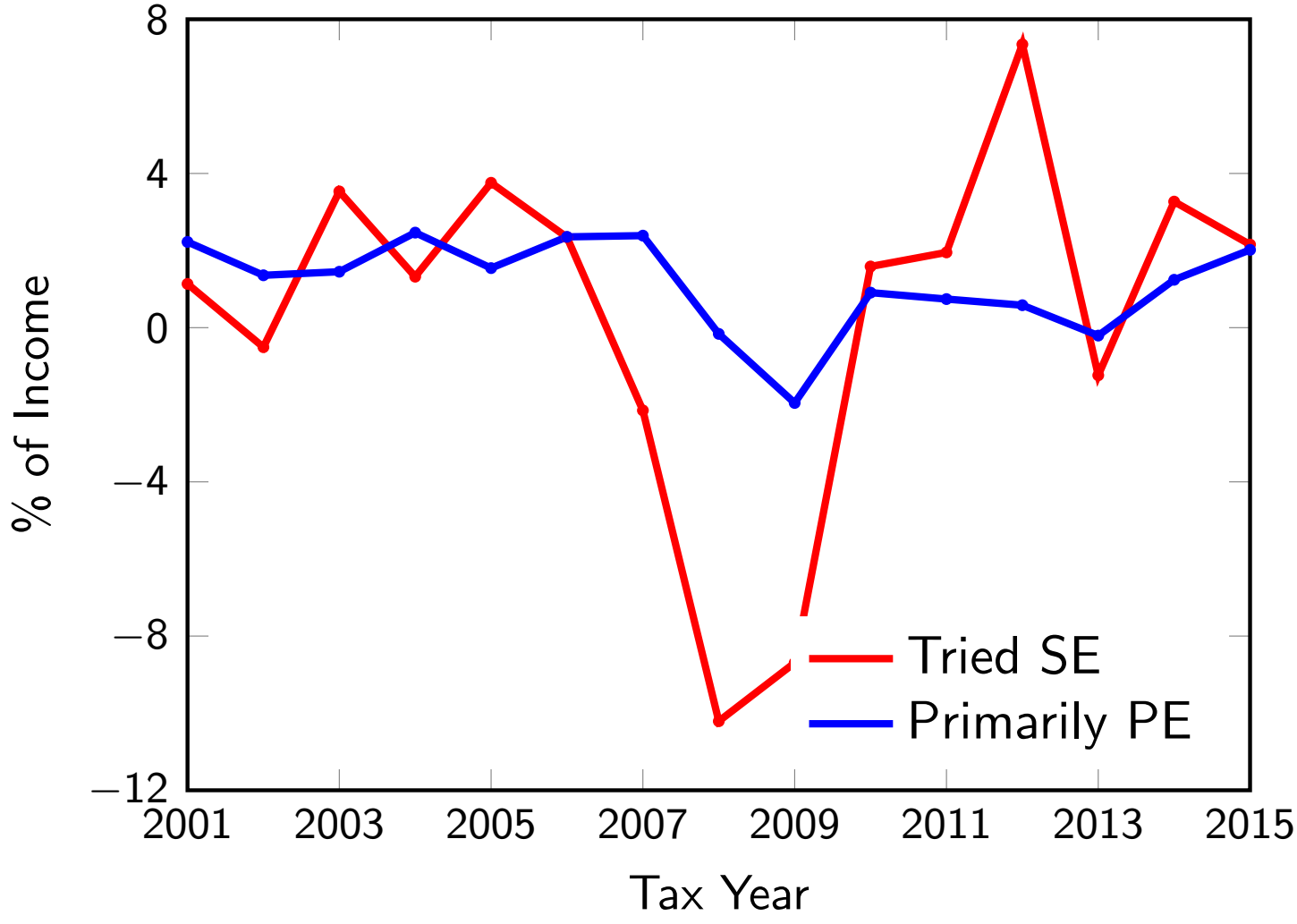


$\Rightarrow$ Easy to rationalize SE choice given estimates of α, $\underline{\Delta c}$

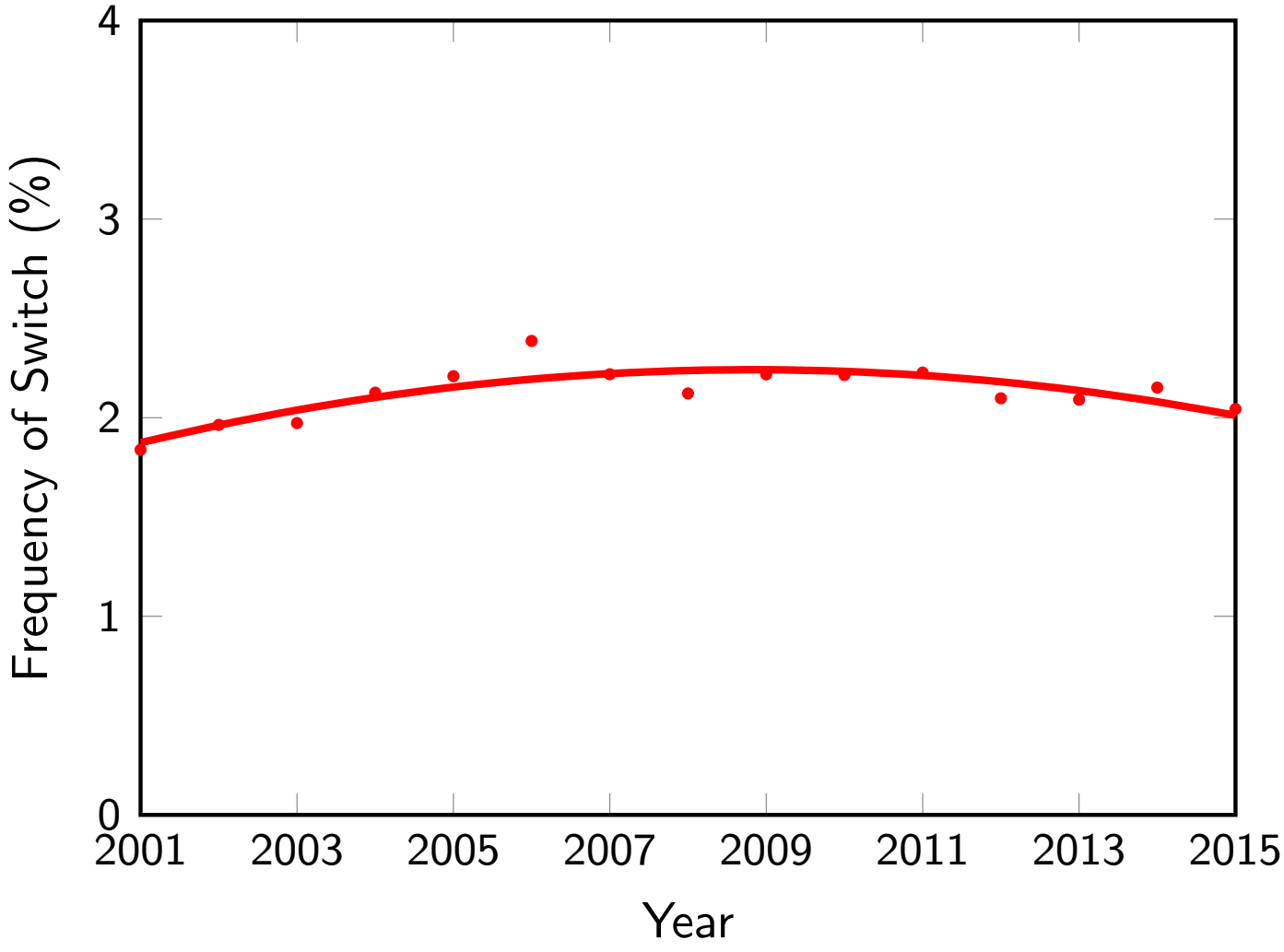
Role for Aggregate Risk

- Hypothesis: business dynamism in decline
- Survey evidence: share of self-employed declining
- IRS evidence:
  - Large decline in estimated time effects in 2008–09
  - But no change in entry or exit rates
  - And no decline in share of SE

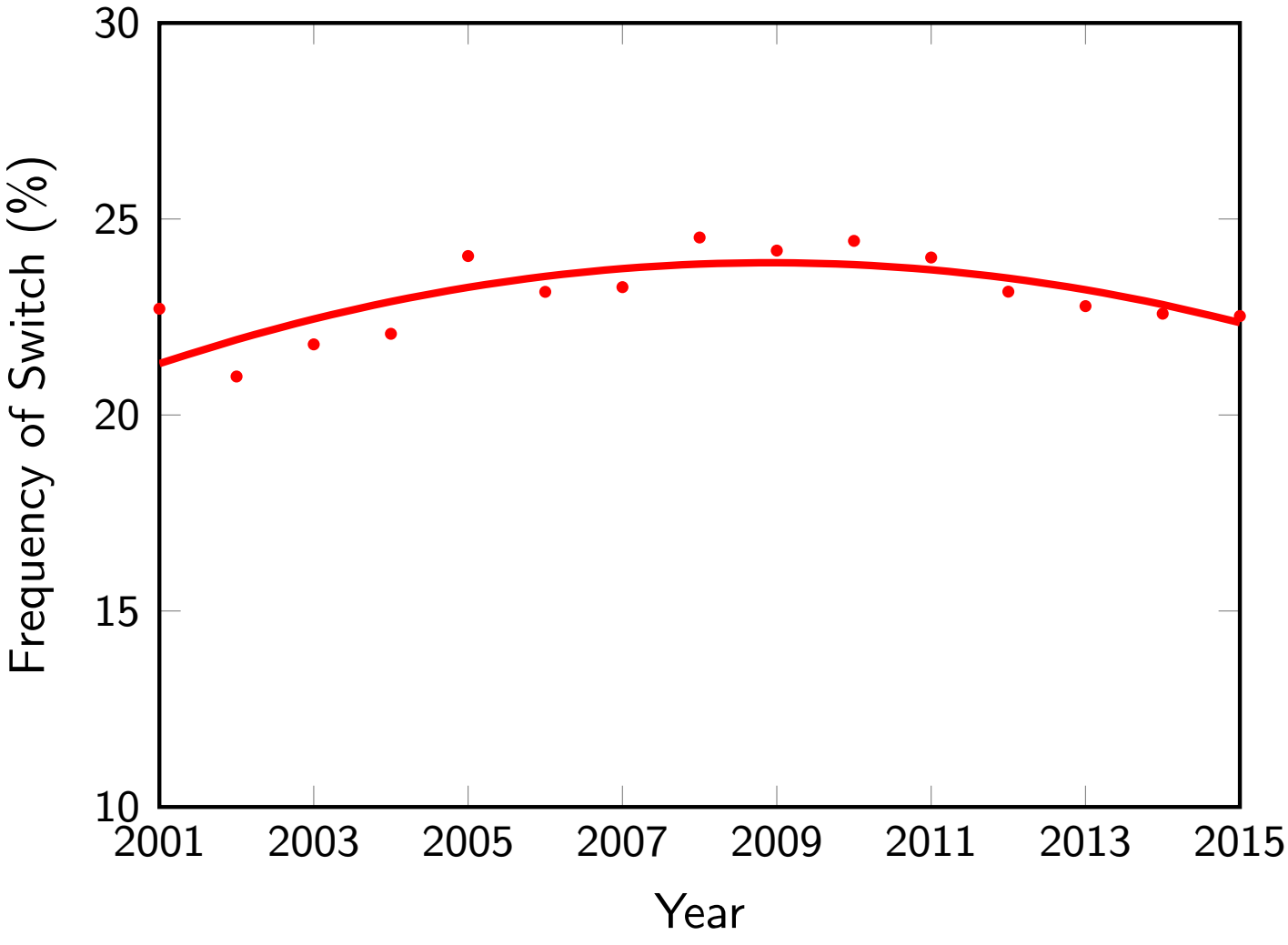
Estimated Time Effects Show Large Decline



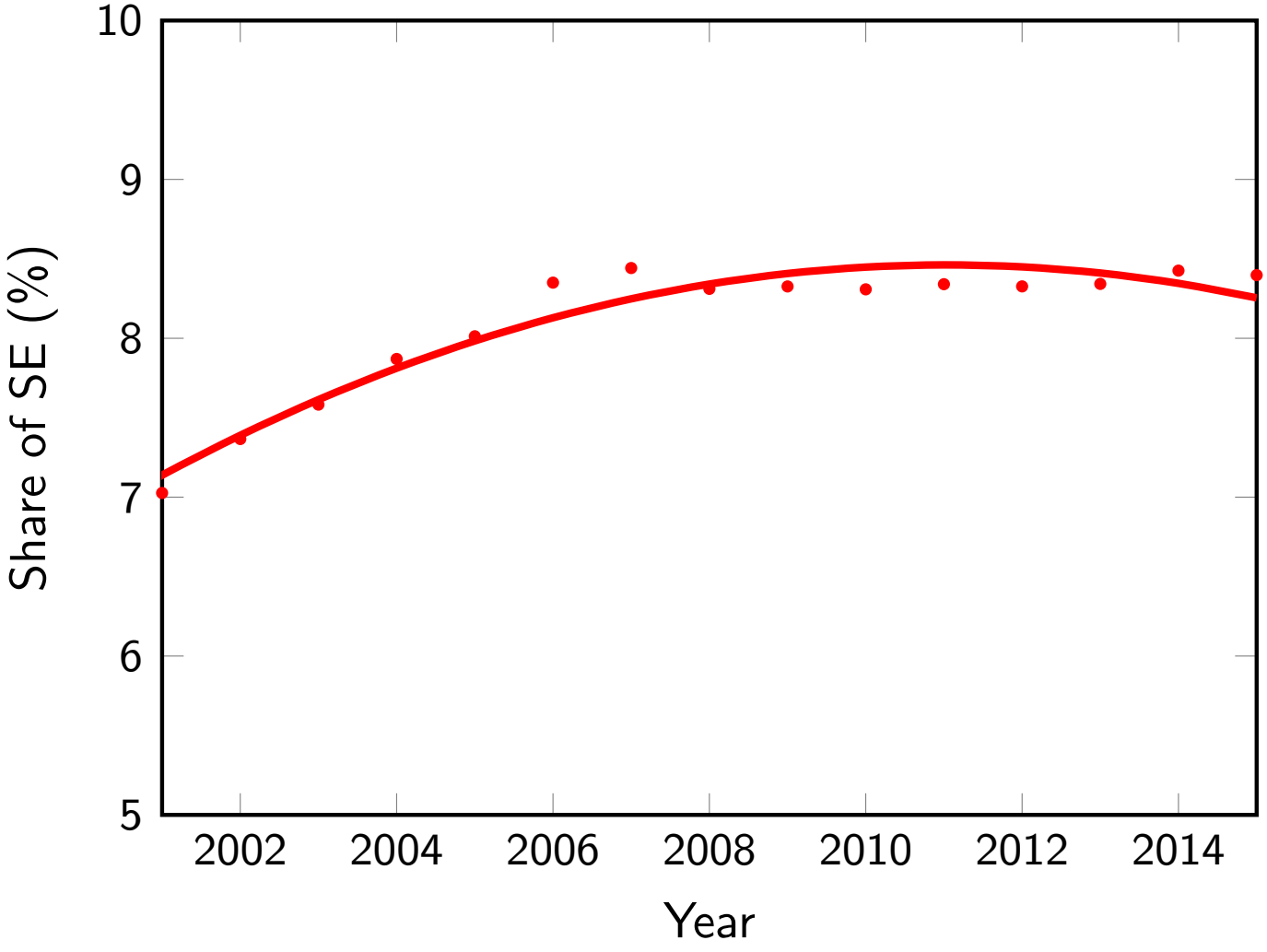
No Change in SE Entry Rates



No Change in SE Exit Rates



No Decline in SE Share



Role for Liquidity Constraints

- Hypothesis: liquidity constraints are binding
- Previous evidence: housing equity increase $\Rightarrow$ entry
- IRS evidence: housing equity increase $\not\Rightarrow$ entry

regression

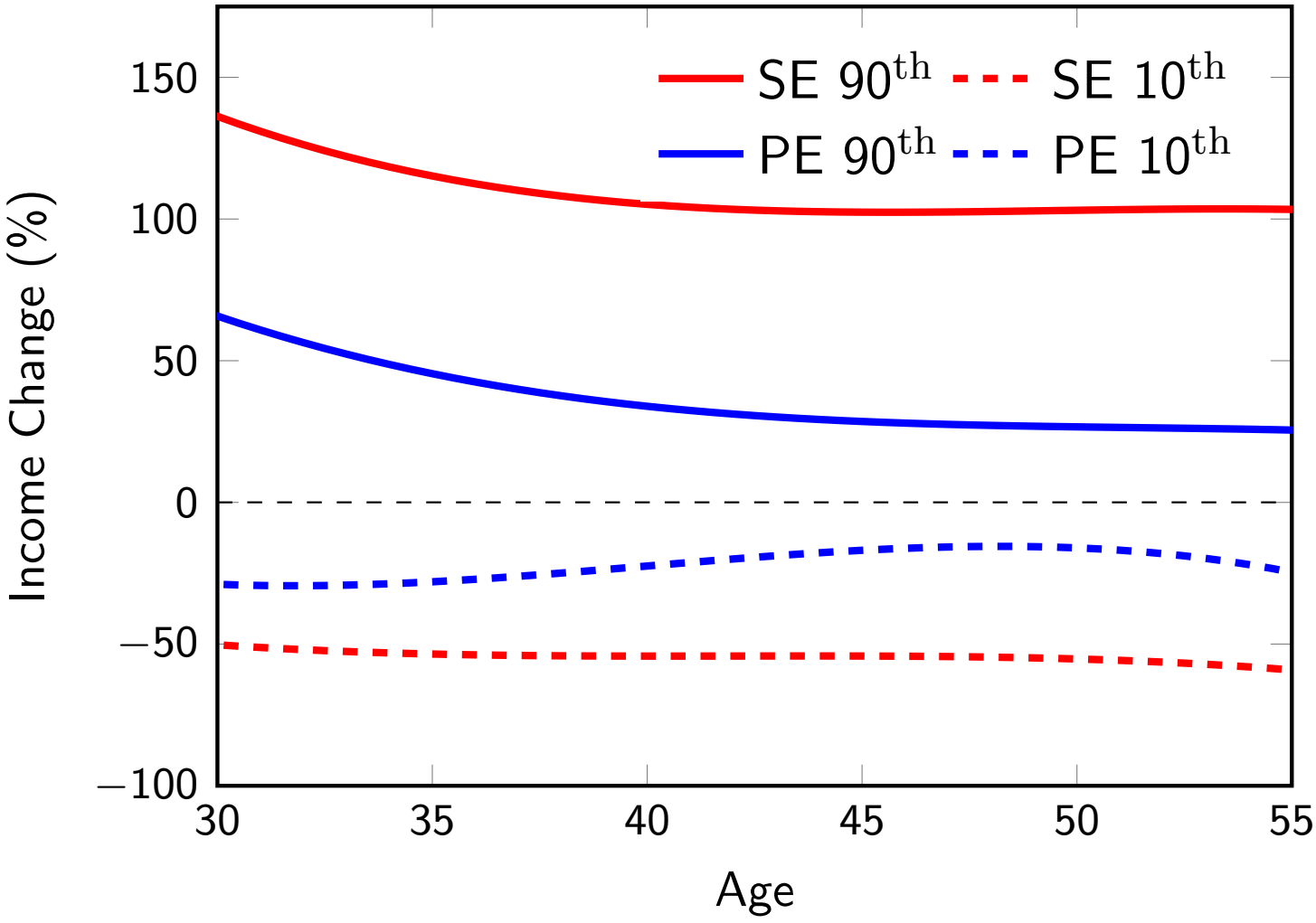
Role for Liquidity Constraints

- Hypothesis: liquidity constraints are binding
- Survey evidence:
  - current entrants have *more* liquid wealth than non-entrants
- IRS evidence: data
  - current entrants have *less* liquid wealth than future entrants

Summary

- Assembled novel longitudinal database for business owners
- Studied patterns of life-cycle income profiles
- Analyzed determinants of entrepreneurial choice
- Provided lots of moments to inform theory, eg,
  - Estimates of life cycle profiles by group
  - Estimates of variability and persistence by group

Evidence of Income Volatility



Evidence of Insurance

back

- Link business and individual filings
- Observe many external resources
  - Business loans and co-owners
  - Individual asset incomes (interest, dividends, capital gains)
  - Spousal incomes (wages, business)
  - Own wages
  - Taxable transfers
- Study founders in first 8 years (Sect. 5.4)
  - Most claimed losses on the business in initial years
  - Almost all had positive individual incomes

Empirical Strategy of Schmalz et al. (2017)

back

$$E_{i,j,t+1} = \beta_0 + \beta_1 \cdot Owner_{i,t} \times \Delta p_{j,t-6\to t-1} + \beta_2 \cdot Owner_{it} + \beta_3 \cdot Owner_{i,t} \times \Delta u_{j,t-6\to t-1} + \beta_4 \cdot \Delta u_{j,t-6\to t-1} + \beta_5 \cdot Z_{i,t} + \beta_6 \cdot Z_{i,t} \times \Delta p_{j,t-6\to t-1} + \delta_{jt} + \epsilon_{i,j,t}$$

where

- $E_{i,j,t+1} = 1$ if switches to SE
- $Owner_{i,t} = 1$ if owns a house in $t-1$
- $\Delta p_{j,t-6\to t-1}$= house-price growth in state j, $t-6$ to $t-1$
- $\Delta u_{j,t-6\to t-1}$= change in unemployment in j, $t-6$ to $t-1$
- $Z_{i,t}$ = control variables
- δ_{jt} = state by year fixed effect

Differences Between Current and Future Switchers

back

Past income type/ Percentile	Ages			
	30	40	50	60
Interest income				
25^{th}	−0.1	−0.4	−1.1	−1.8
50^{th}	−0.0	−0.1	−0.3	−0.4
75^{th}	0.0	−0.0	−0.0	−0.0
Dividend income				
25^{th}	−0.1	−0.3	−0.9	−1.6
50^{th}	−0.0	−0.1	−0.2	−0.3
75^{th}	0.0	0.0	−0.0	0.0
Capital gains				
25^{th}	−0.2	−0.9	−3.6	−6.3
50^{th}	−0.0	−0.1	−0.5	−1.1
75^{th}	0.0	0.0	0.0	0.2