On the Nature of Entrepreneurship

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

[1]Department of Economics
University of Minnesota

[2]Office of Tax Analysis
Department of Treasury

[3]Internal Revenue Service
Department of Treasury

Disclaimer

The authors thank Anne Parker and Barry Johnson for facilitating this project through the Joint Statistical Research Program of the Statistics of Income Division of the United States Internal Revenue Service. May and McGrattan are IRS employees without pay under an agreement made possible by the Intragovernmental Personnel Act of 1970 (5 U.S.C. 3371-3376). Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Internal Revenue Service or the U.S. Department of the Treasury, or the National Science Foundation. All results have been reviewed to ensure that no confidential information is disclosed. All data work for this project involving confidential taxpayer information was done at IRS facilities, on IRS computers, by IRS employees, and at no time was confidential taxpayer data ever outside of the IRS computing environment.

This Paper

- Informs theories of entrepreneurship
- How?
  - Assembles novel longitudinal database of business owners
  - Studies patterns of life-cycle income profiles
  - Analyzes determinants of entrepreneurial choice

Most Previous Work

- Uses surveys with
  - Top-coding
  - Short panels
- Concludes that self-employed (relative to peers)
  - Have flatter life-cycle profiles
  - Enter self-employment with lower past labor income
  - Enter with higher past asset income
- Motivates theories where entrepreneurs
  - Earn large non-pecuniary benefits
  - Are misfits
  - Face liquidity constraints

In Contrast to Literature

- Use administrative data with

  ▸ Even in cross-section, IRS $\neq$ CPS

  - No Top-coding
  - Long panels
- Conclude that self-employed (relative to peers)
  - Have significantly steeper life-cycle profiles
  - Enter self-employment with higher past labor income
  - Enter with lower past asset income
- Motivate theories where entrepreneurs
  - Make significant investments in business
  - Are not misfits
  - Face few liquidity constraints

Data

Sample

- Primary source: administrative IRS data
  - Balanced panel of living individuals with US SSN
  - Tax years 2000-2015
  - Birth cohorts 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

Employment Status

- Self-employed (SE) in a given year if:
  - |SE income| $>$ 5,000 in 2012\$ **and** at least one of:
    - |SE income| $>$ PE income or
    - Share of gross profits $>$ PE income or
    - Share in business $\times$ employees ≥ 1
- Paid-employed (PE) in a given year if:
  - Not SE
  - PE income $>$ 5,000 in 2012\$
- Non-employed (NE) in a given year if:
  - Not SE or PE

Skill and Education Measures

Skills:

- Individuals with occupation in e-filing
  - Map entry to SOC code
  - Map SOC to cognitive, interpersonal, and manual skills (as in Lise and Postel-Vinay 2020)

- Individuals with missing codes
  - Use AI tools and data for peers with codes

Education:

- Use CPS-based classifier

Life-Cycle Profile Estimation

Object of Interest

Income(Age | Individual and aggregate factors)

Estimation Procedure

- Statistical model for income:

$$y_{it} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a(i,t)} \gamma^a_{c(i),g(i)} + \epsilon_{i,t}$$

where

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

Estimation Procedure

- Statistical model for income:

$$y_{it} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a(i,t)} \gamma^a_{c(i),g(i)} + \epsilon_{i,t}$$

fixed effects

where

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

Estimation Procedure

- Statistical model for income:

$$y_{it} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a(i,t)} \gamma^a_{c(i),g(i)} + \epsilon_{i,t}$$

time effects

where

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

Estimation Procedure

- Statistical model for income:

$$y_{it} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a(i,t)} \gamma^a_{c(i),g(i)} + \epsilon_{i,t}$$

age effects

where

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

Estimation Procedure

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

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

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

▸ More details on identification assumptions

Application: set $\mathcal{G}$ with 46,080 subgroups

- Time-invariant characteristics include usual ones:
  - Cohort, gender, educated, skilled (cognitively, interpersonally, manually), industry, married, children
- Plus partition sample based on *Employment attachment*
  - Attached SE, Attached PE, Switchers ▸ Definitions

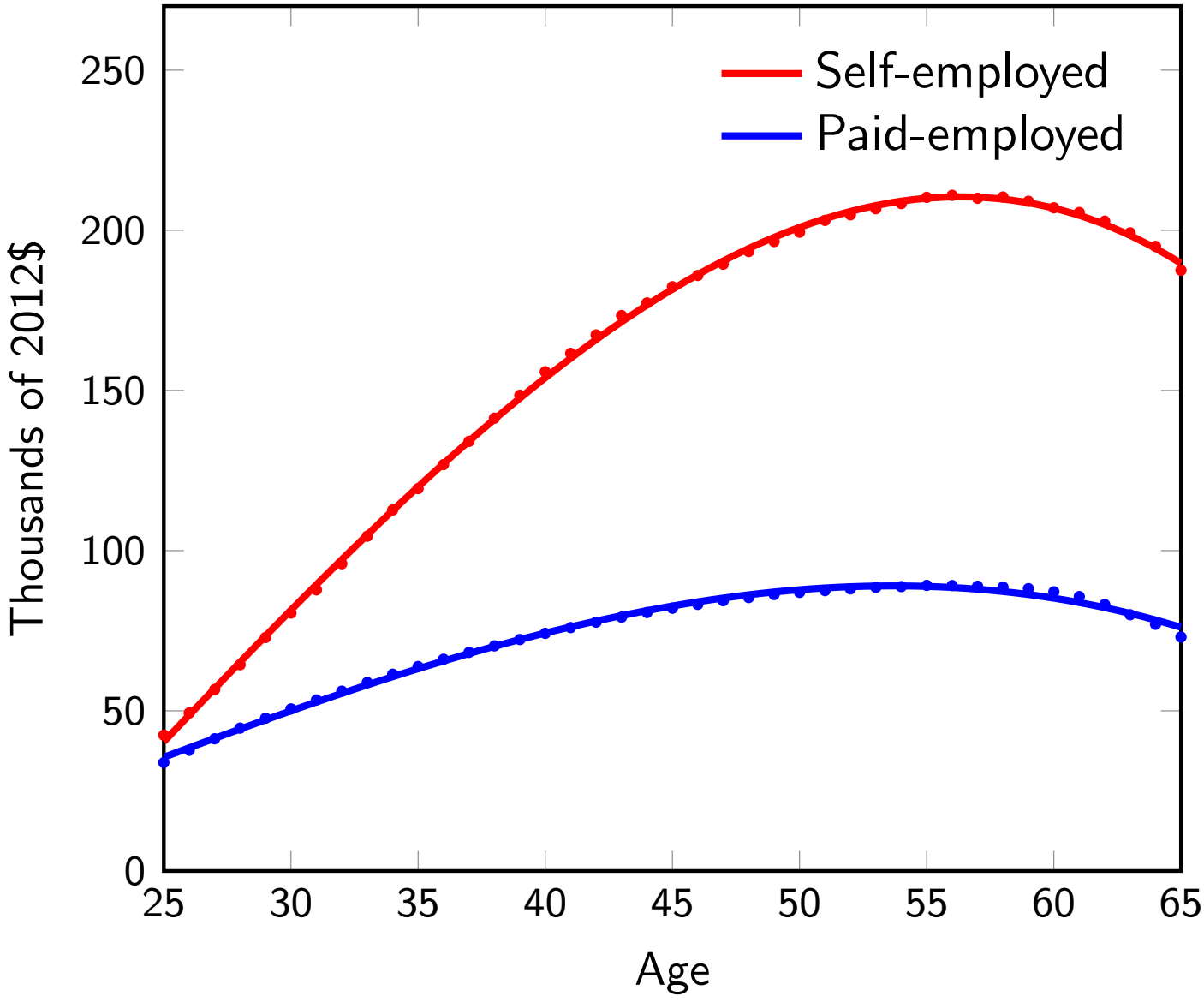
Main Empirical Results

Income and Growth Profiles

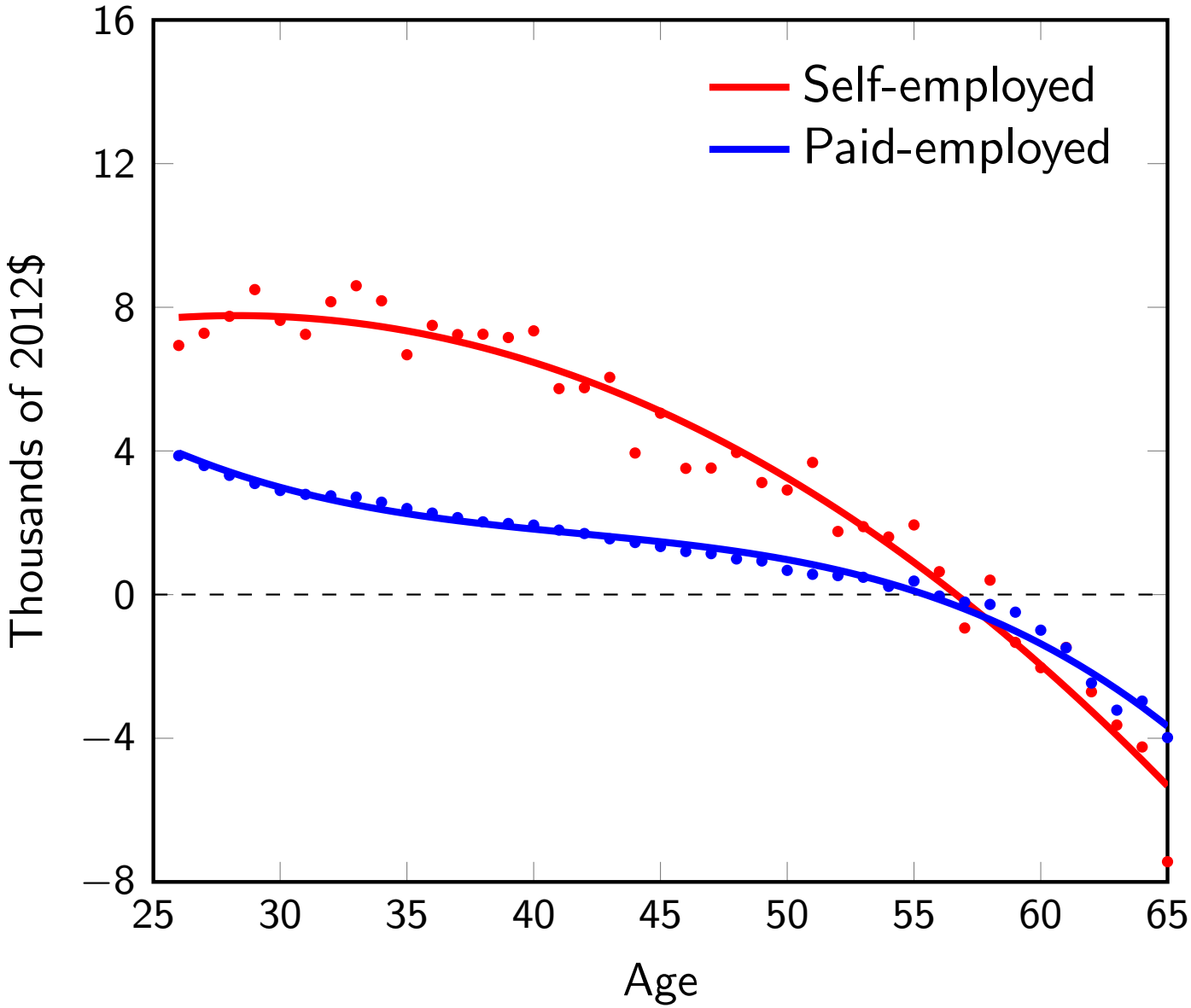
- Attached self-employed
  - Income similar on average to paid-employed when 25
  - Growth significantly higher and more persistent

$\Rightarrow$ Entrepreneurial investment does pay

Income Profiles: Attached Subsamples



Growth Profiles: Attached Subsamples



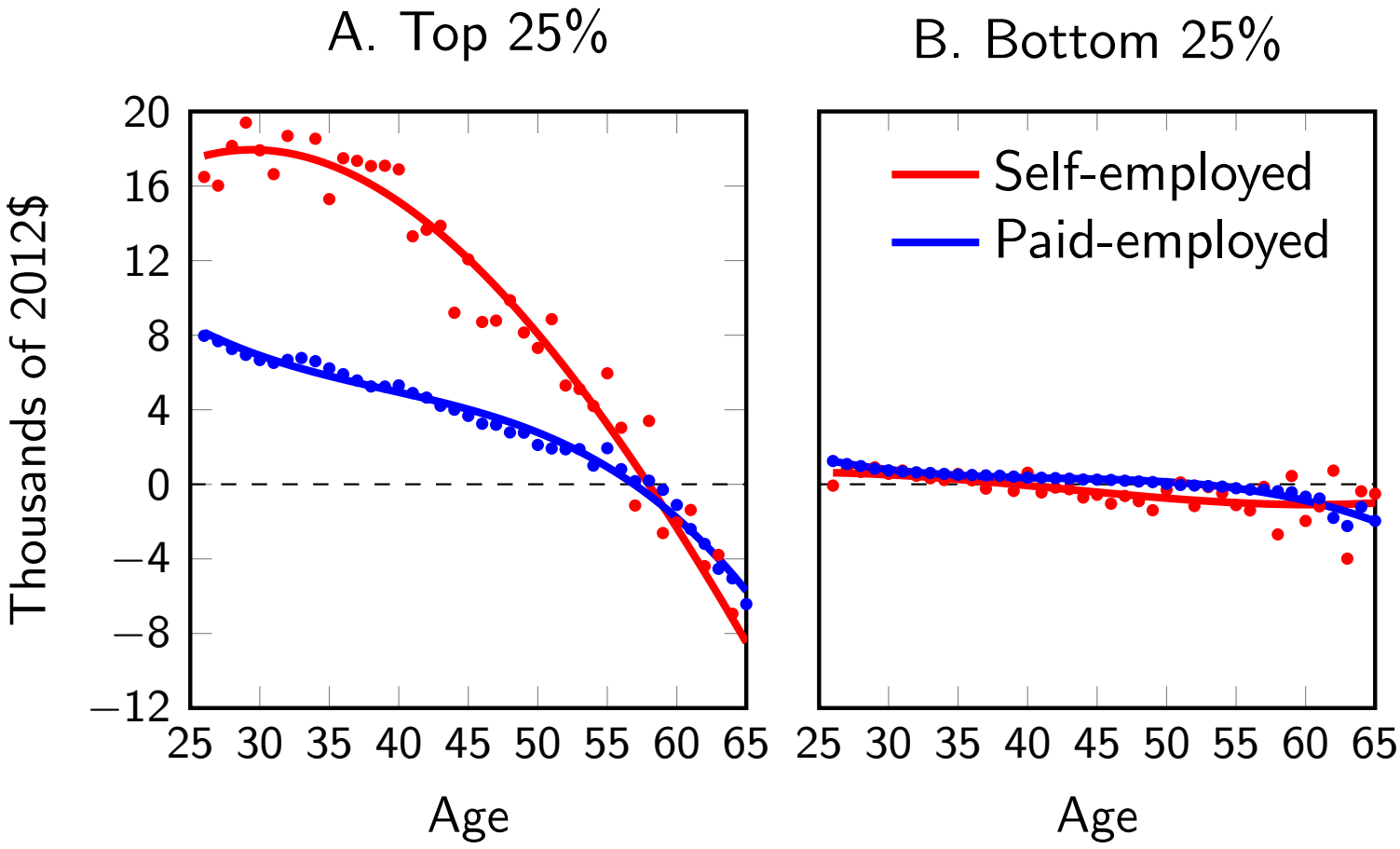
▸ Time Effects

Income and Growth Profiles

- Attached self-employed
  - Income similar on average to paid-employed when 25
  - Growth significantly higher and more persistent

$\Rightarrow$ But there are large differences for top/bottom 25% earners

Growth Profiles: Income Ranks



Disaggregating: An Example

Disaggregating: An Example

- Consider

Disaggregating: An Example

- Consider
  - Men

Disaggregating: An Example

- Consider
  - Men
  - Married

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled
  - Interpersonally skilled

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled
  - Interpersonally skilled
  - Not manually skilled

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled
  - Interpersonally skilled
  - Not manually skilled
  - Working in professional services

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled
  - Interpersonally skilled
  - Not manually skilled
  - Working in professional services
  - Attached to paid- or self-employment

Disaggregating: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled
  - Interpersonally skilled
  - Not manually skilled
  - Working in professional services
  - Attached to paid- or self-employment

$\Rightarrow$ Just 2 of the 46,080 groups

Growth Profiles: Disaggregated Group

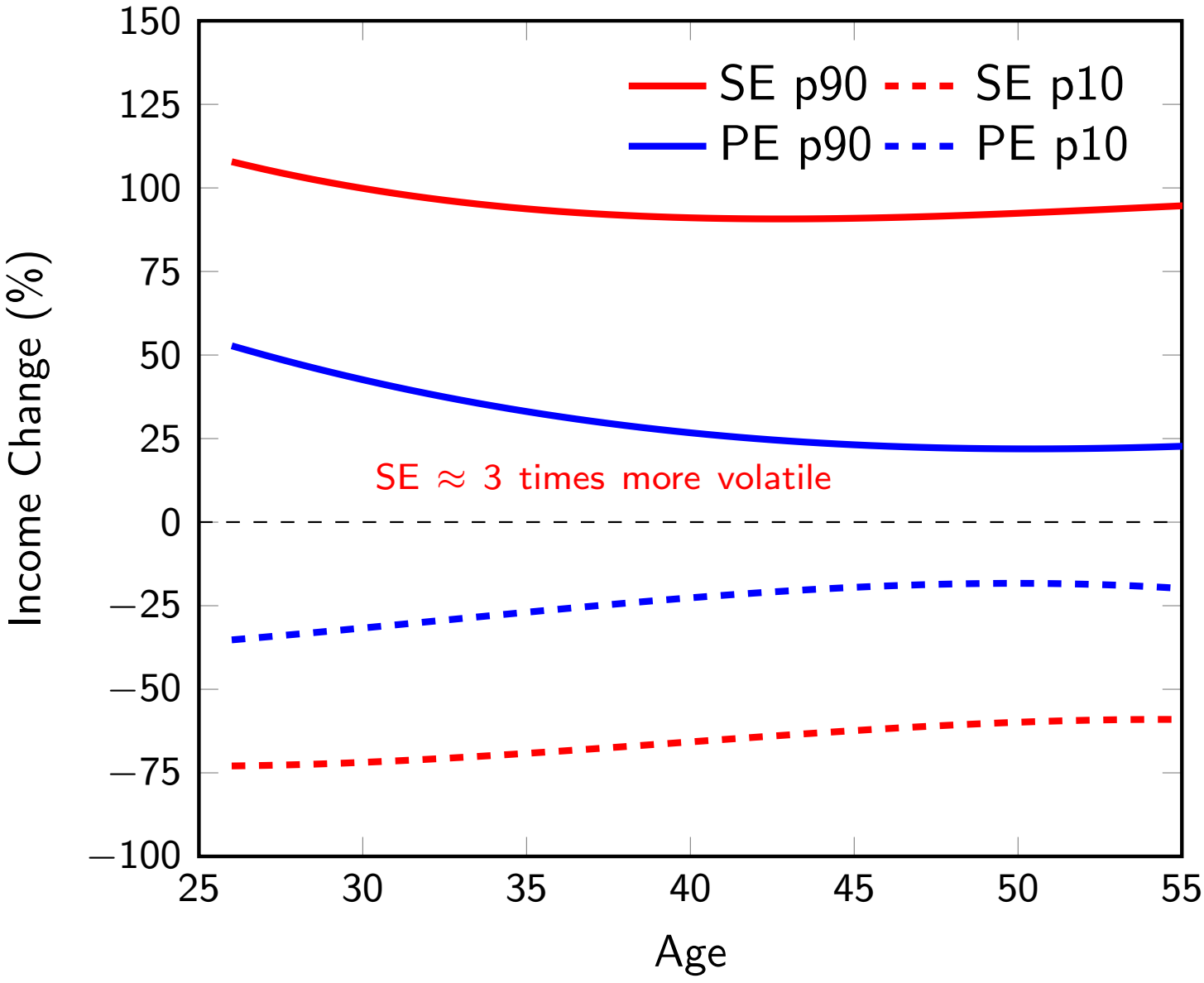
Volatility Patterns

- Large literature on risk in entreprenurship
  - Is SE more risky than PE? By how much?
  - Are differences in growth driven by increasing risk over age?
- Compute distribution of residuals (net of time-age effects)

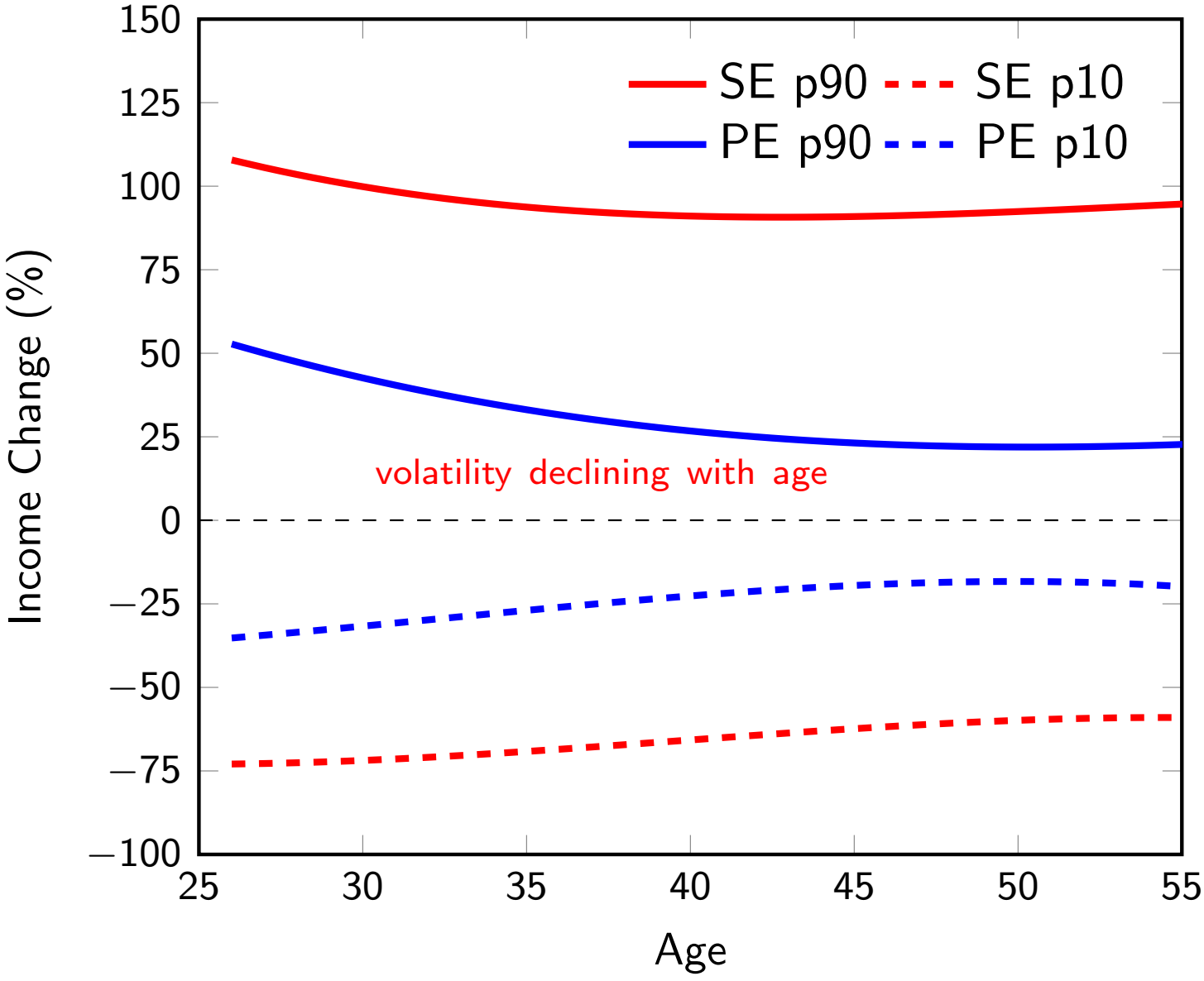
$$\Delta \epsilon_{i,a} / |y_{i,a-1}|$$

- Compare SE and PE
  - Plot 10^{th} and 90^{th} percentiles by age and employment status

Income More Volatile for Attached SE



Income More Volatile for Attached SE



Back of the Envelope Welfare Calculation

- With assumptions on
  - Preferences, eg, Epstein-Zin with $\rho \to 0$

$$V_t(\{c_j\}_{j=t}^{\infty}) = \left[(1-\beta)c_t^{\rho} + \beta(E_t V_{t+1}^{\alpha})^{\rho/\alpha}\right]^{\frac{1}{\rho}}$$

  - Income processes, eg, random walk r_t plus temporary z_t

- Can match moments for income growth:
  - 90-10 difference in growth, $Q = 2.56\sqrt{\sigma_r^2 + 2\sigma_z^2}$
  - Autocorrelation, $A = -\sigma_z^2/(\sigma_r^2 + 2\sigma_z^2)$

- To infer fraction of wealth λ sacrificed to fully insure $c = y$

$$\lambda = -0.5\alpha\beta\sigma_r^2$$

Back of the Envelope Welfare Calculation (SE/PE Ratio)

- With assumptions on
  - Preferences, eg, Epstein-Zin with $\rho \to 0$

$$V_t(\{c_j\}_{j=t}^{\infty}) = \left[(1-\beta)c_t^{\rho} + \beta(E_t V_{t+1}^{\alpha})^{\rho/\alpha}\right]^{\frac{1}{\rho}}$$

  - Income processes, eg, random walk r_t plus temporary z_t

- Can match moments for income growth:
  - 90-10 difference in growth, $Q = 2.56\sqrt{\sigma_r^2 + 2\sigma_z^2}$ (≈ 3)
  - Autocorrelation, $A = -\sigma_z^2/(\sigma_r^2 + 2\sigma_z^2)$ (≈ 1)

- To infer fraction of wealth λ sacrificed to fully insure $c = y$

$$\lambda = -0.5\alpha\beta\sigma_r^2 \qquad (\approx Q^2 = 9)$$

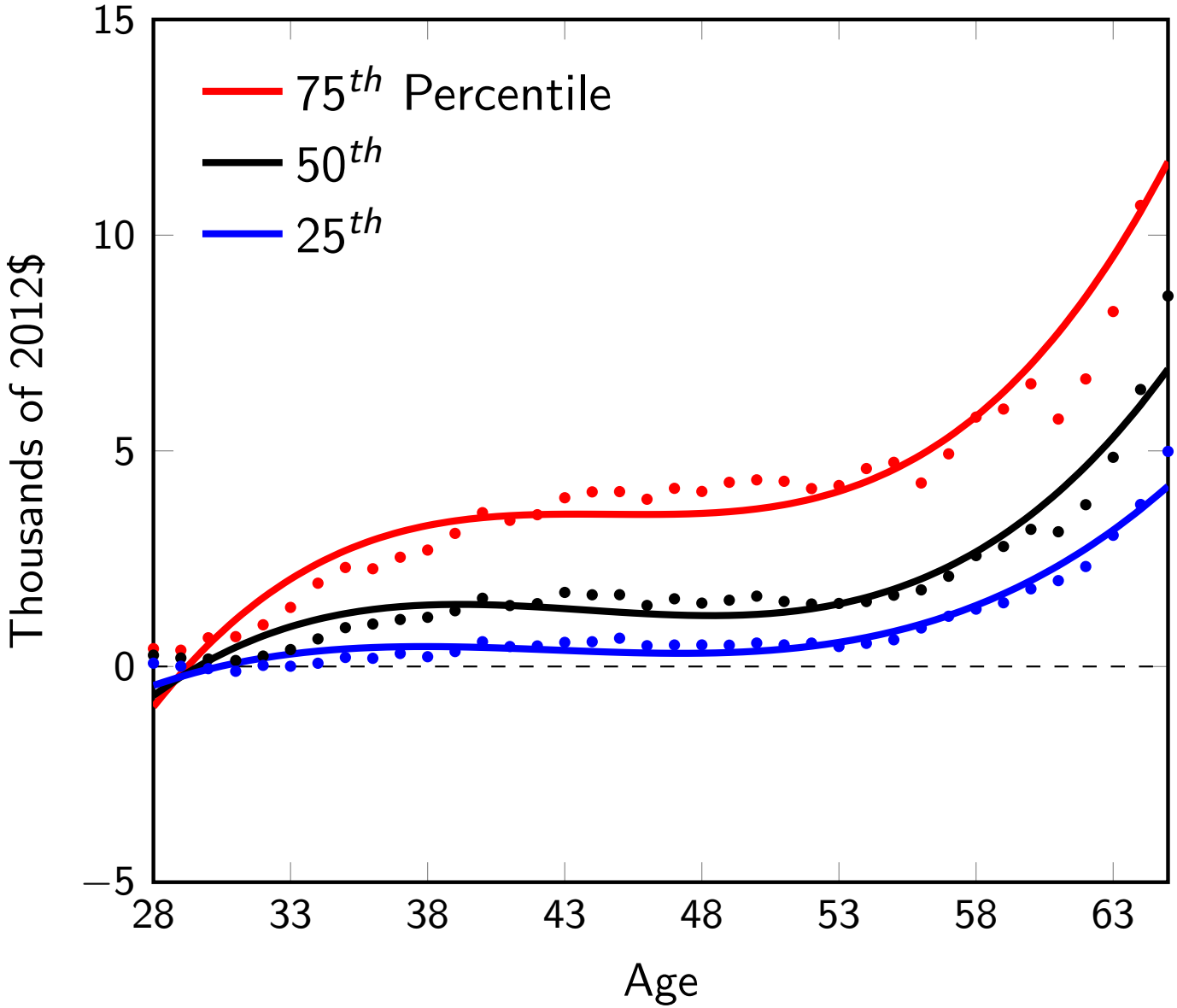
Entrepreneurial Choice

- Entry and exit rates
  - Results similar to surveys
- Use switchers to study
  - Key determinants of choosing self-employment

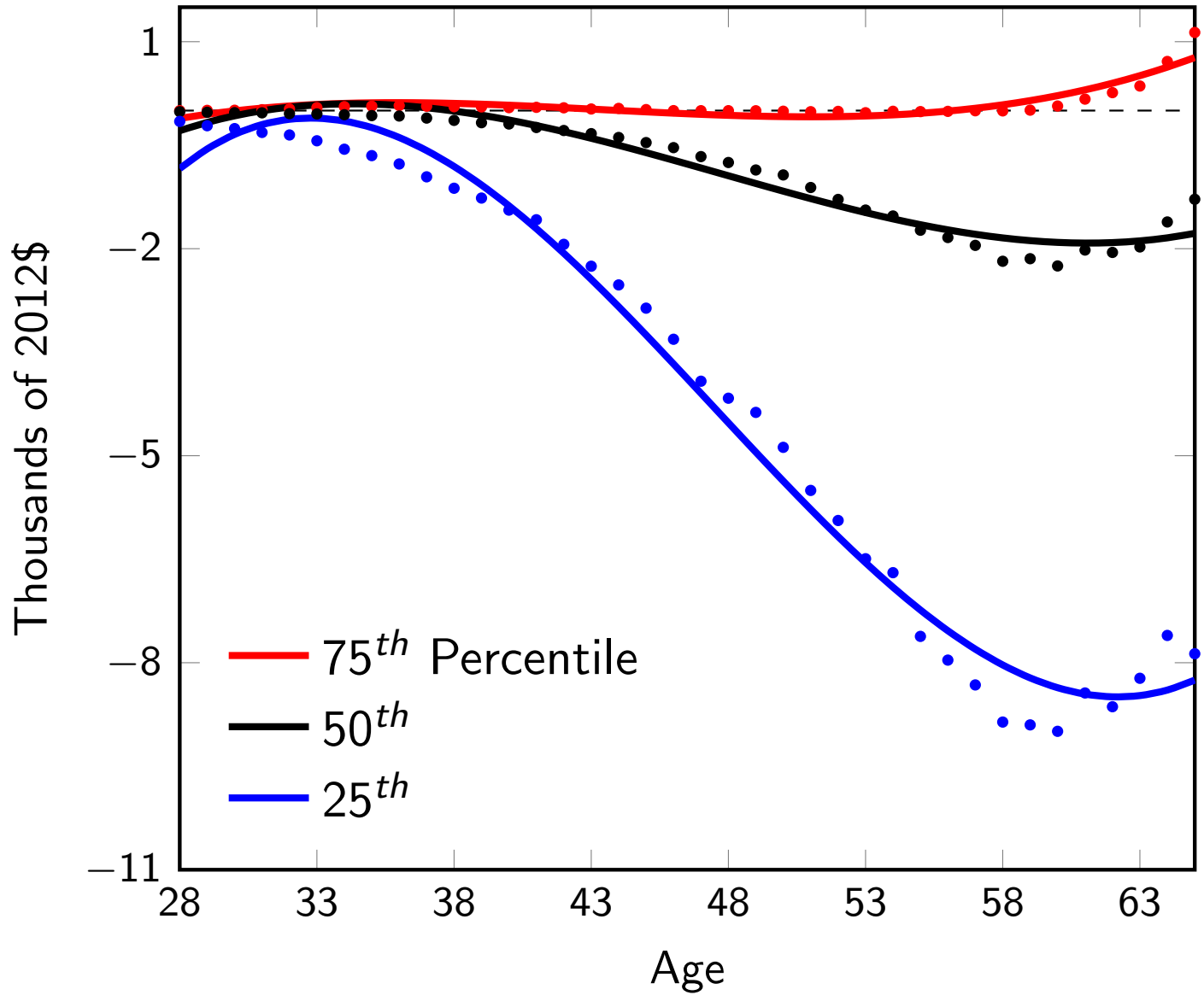
Determinants of Self-Employment

- Compare SE entrants to "similar" peers
  - One-time entrants into SE ("Treatment")
  - Future switchers with same characteristics ("Control")
- Assess "misfit" hypothesis for SE
  - Compare wage income before entry
- Assess "financial-friction" hypothesis for SE
  - Compare asset income before entry

Past Wage Incomes Higher for Switchers



Past Asset Incomes Lower for Switchers



Start-ups: Income in Initial Years

- Consider S-corp/partnership founders in 1970-75 cohort
  - First Schedule K-1 in year business starts
  - Eight years of consecutive tax filings

- Year: business/owner has negative income (%)

  1: 45 / 10
  2: 35 / 9
  3: 32 / 8

- Year: business/owner income first positive (%)

  1: 53 / 90
  2: 19 / 5
  3: 8 / 2

Informing Theory

Empirically-Motivated Features

- Patterns in the data
  - Hump-shaped and persistent income growth
  - Declining exit rates
  - Volatility decreasing with age
- Empirical results suggest three model features
  - Investment in self-created intangible assets ▸ Evidence
  - Incomplete information about entrepreneurial productivity
  - Slow adjustment in achieving optimal size

Modeling Intangibles

- State vector $s = [a, \kappa, j, \epsilon, z, \mu]$
- Dynamic program for entrepreneur

$$
\begin{aligned}
V_k(s) &= \max\{U(c, \ell) + \beta EV(s')\} \\
a' &= (1 + r)a + pe^z f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0 \\
\kappa' &= (1 - \delta_\kappa)\kappa + f_\kappa(h_\kappa, e) \\
\ell &= 1 - h_y - h_\kappa
\end{aligned}
$$

- Two production technologies:
  - $f_y(\kappa, h_y, k, n)$: goods and services
  - $f_\kappa(h_\kappa, e)$: new intangible assets

Modeling Intangibles

- State vector $s = [a, \kappa, j, \epsilon, z, \mu]$

  financial assets

- Dynamic program for entrepreneur

$$
\begin{aligned}
V_k(s) &= \max\{U(c,\ell) + \beta EV(s')\} \\
a' &= (1+r)a + pe^z f_y(\kappa, h_y, k, n) - (r+\delta_k)k - wn - e - c \geq 0 \\
\kappa' &= (1-\delta_\kappa)\kappa + f_\kappa(h_\kappa, e) \\
\ell &= 1 - h_y - h_\kappa
\end{aligned}
$$

- Two production technologies:
  - $f_y(\kappa, h_y, k, n)$: goods and services
  - $f_\kappa(h_\kappa, e)$: new intangible assets

Modeling Intangibles

- State vector $s = [a, \kappa, j, \epsilon, z, \mu]$

  intangible assets

- Dynamic program for entrepreneur

$$
\begin{aligned}
V_k(s) &= \max\{U(c, \ell) + \beta EV(s')\} \\
a' &= (1 + r)a + pe^z f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0 \\
\kappa' &= (1 - \delta_\kappa)\kappa + f_\kappa(h_\kappa, e) \\
\ell &= 1 - h_y - h_\kappa
\end{aligned}
$$

- Two production technologies:
  - $f_y(\kappa, h_y, k, n)$: goods and services
  - $f_\kappa(h_\kappa, e)$: new intangible assets

Modeling Intangibles

- State vector $s = [a, \kappa, j, \epsilon, z, \mu]$

  age

- Dynamic program for entrepreneur

$$
\begin{aligned}
V_k(s) &= \max\{U(c, \ell) + \beta EV(s')\} \\
a' &= (1 + r)a + pe^z f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0 \\
\kappa' &= (1 - \delta_\kappa)\kappa + f_\kappa(h_\kappa, e) \\
\ell &= 1 - h_y - h_\kappa
\end{aligned}
$$

- Two production technologies:
  - $f_y(\kappa, h_y, k, n)$: goods and services
  - $f_\kappa(h_\kappa, e)$: new intangible assets

Modeling Intangibles

- State vector $s = [a, \kappa, j, \epsilon, z, \mu]$

  true and predicted skills

- Dynamic program for entrepreneur

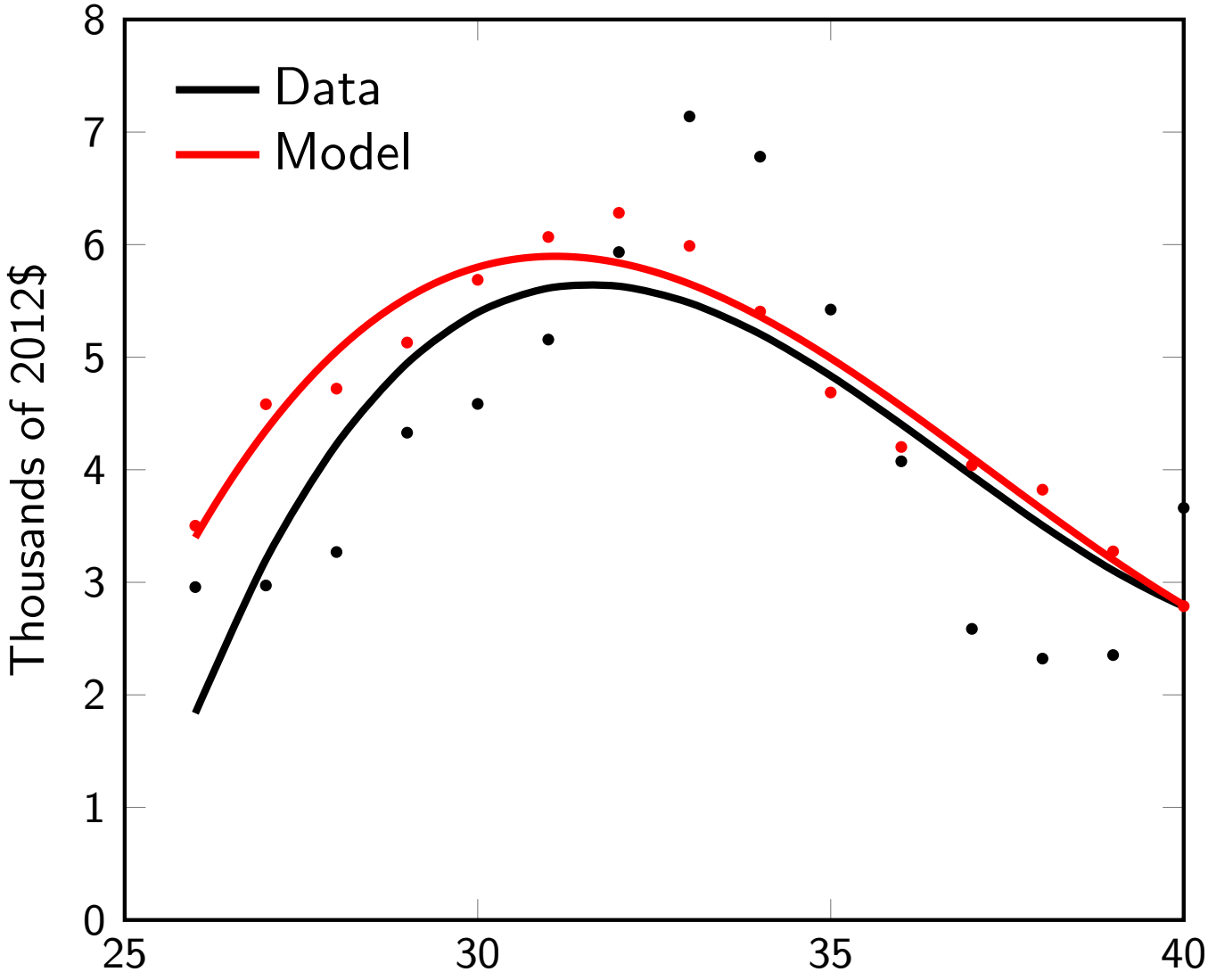
$$
\begin{aligned}
V_k(s) &= \max\{U(c, \ell) + \beta EV(s')\} \\
a' &= (1 + r)a + pe^z f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0 \\
\kappa' &= (1 - \delta_\kappa)\kappa + f_\kappa(h_\kappa, e) \\
\ell &= 1 - h_y - h_\kappa
\end{aligned}
$$

- Two production technologies:
  - $f_y(\kappa, h_y, k, n)$: goods and services
  - $f_\kappa(h_\kappa, e)$: new intangible assets

Comparing Growth Profiles

- Choose income shocks consistent with IRS micro data
- Simulate time series over the life cycle
- Aggregate simulations using IRS counts and entry ages
- Construct growth differential for self-employed:
  - Stayers: attached to self-employment past age 35
  - Switchers: ran a business at least 5 years but exited by 35

Growth Differentials for Young Entrepreneurs



Conclusion

- Assembled novel longitudinal database for business owners
- Estimated life-cycle income profiles for many groups
- Developed prototype model of entrepreneurs
- Studied model predictions for IRS data

Appendix

Identification

- Two identifying assumptions
  - Age effects are same across binned cohorts (≥ 2)
  - Average time effect satisfies (where $\overline{y}_{g,t_0}$ is avg income for g):

$$\frac{\overline{\Delta\beta_g}}{\overline{y}_{g,t_0}} = \frac{\mu_g}{T}\sum_t (1+\mu_g)^t$$

- Allows flexibility when set $\mathcal{G}$ large

◂ Back

Employment Attachment

- *Attached* (SE or PE) if:
  - Fewer than 2 switches in status during sample
  - No itermediate spells of non-employment
- *Mostly switchers* if:
  - In SE or PE for 12+ years
  - No intermediate spells of non-employment
- *Any non-employment* if:
  - Switched in/out of NE from SE or PE at least once
  - Or, 5 years of NE during sample

Back

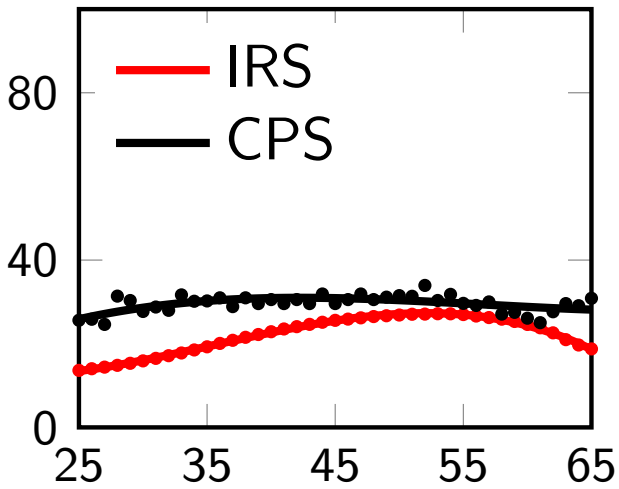
Evidence of Business Intangibles

- Business sale is taxable event for buyer and seller
- Forms 8594, 8883 show assets primarily intangible, eg
  - Customer bases, client lists, non-compete covenants
  - Licenses, permits, trademarks, tradenames
  - Workforce in place
  - Goodwill and on-going concern value

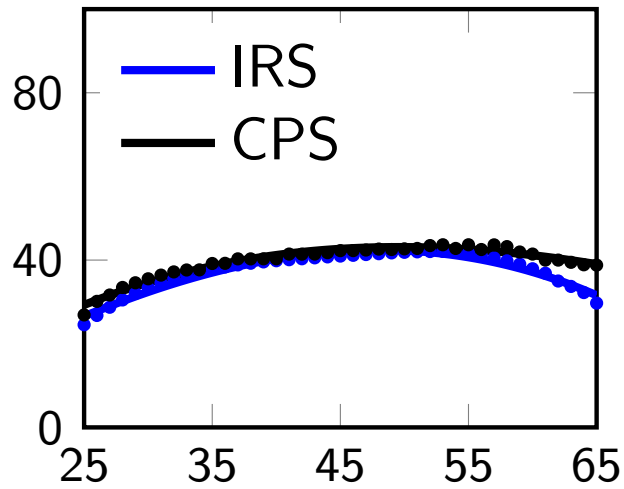
◂ Back

Empirical Moments: IRS vs CPS (Thous. 2012$)

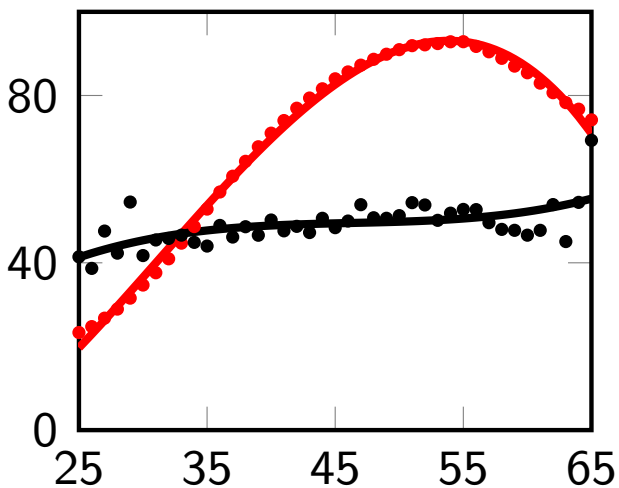
SE Median Income



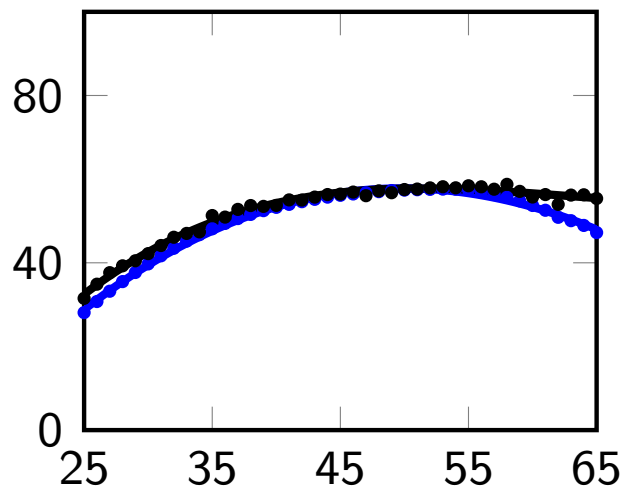
PE Median Income



SE Mean Income



PE Mean Income



◂ Back

Time Effects Relative to Income

Back