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

- Assembles novel longitudinal database of business owners
- Estimates life-cycle income profiles for 35,000 groups
- Compares profiles for similar self- and paid-employed
  - Growth and volatility patterns
  - Determinants of entrepreneurial choice

Motivation

- Provides updated answers to:
  - Does entrepreneurship pay?
  - Is there scope for shrinking the tax gap?
- Results inform:
  - Entrepreneurial theories
  - Tax administration

Preview of Findings

- Provides updated answers to:
  - Does entrepreneurship pay? Yes
  - Is there scope for shrinking the tax gap? Yes
- Results inform:
  - Entrepreneurial theories
  - Tax administration

Most Previous Work

- Uses surveys with
  - Top-coding
  - Short panels
- Concludes that SE (relative to peers)
  - Flatter life-cycle profiles
  - Enter SE 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
  - No Top-coding
  - Long panels
- Conclude that SE (relative to peers)
  - Have significantly steeper life-cycle profiles
  - Enter SE with higher past labor income
  - Enter with lower past asset income
- Motivate theories where entrepreneurs
  - Make significant investments in business
  - Experiment to learn entrepreneurial productivity
  - Face few liquidity constraints

Today's talk

- Data
  - Sample
  - Income measures
  - Skill and education imputations
  - Cross-sectional comparisons with CPS
- Life-cycle profile estimation
  - Potential challenges
  - Econometric approach
  - Income and growth profiles by group
- Entrepreneurial choice
  - Entry and exit
  - Characteristics of entrants
- Theoretical predictions

Data

Sample

- Primary source: administrative IRS data
  - Balanced panel of living individuals with US SSN
  - Birth cohorts 1950-1975
  - Available 1996-2015
- Merge in: Schedule C and K-1 data
  - Owners of pass-through businesses
  - Available 2000-present

Income Measures

- Self-employment (SE) income
  - Schedule C net profit of sole proprietors
  - Schedule K-1 ordinary business income of
    - Individual partners
    - S-corporation owners
  - W-2 wages of S-corporation owners
- Paid-employment (PE) income
  - W-2 wages of non-owner employees

Employment Status

- Self-employed (SE) in a given year if:
    - $|\text{SE income}| > 5{,}000$ in 2012\$
      and at least one:
        - $|\text{SE income}| > |\text{PE income}|$
        - Share in business $\times$ employees ≥ 1
        - Share of gross profits $>$ PE income

- Paid-employed (PE) in a given year if:
    - Not SE
    - W-2 earnings $> 5{,}000$ 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

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

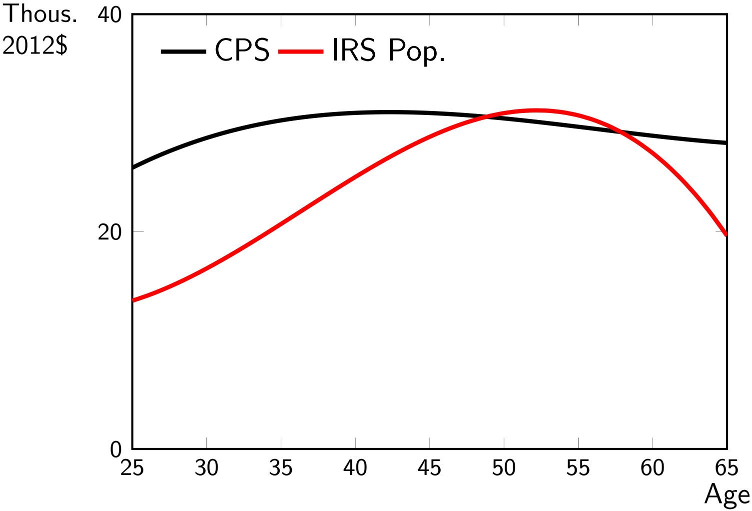
Education:

- Individuals classified as college-educated if
  - Filed 1098-T
  - Listed occupation as student
  - Predicted as so by CPS-based classifier

Empirical Moments: IRS vs CPS

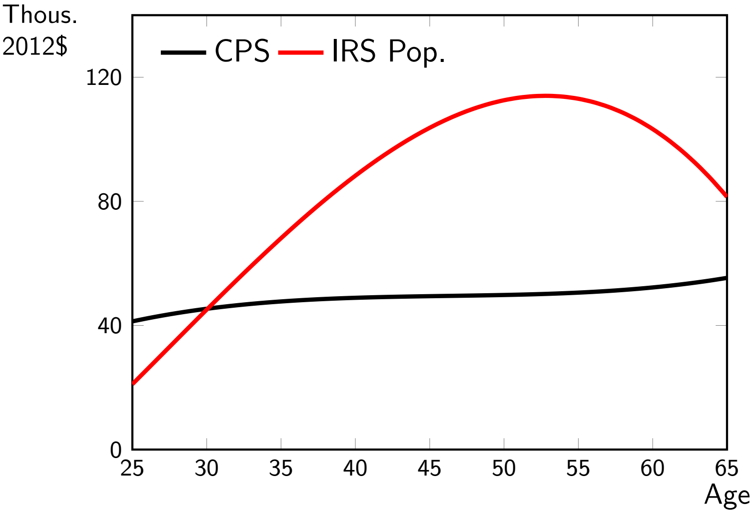
- Use only two criteria for SE assignment
  - $|\text{SE income}| > 5{,}000$ in 2012\$ and
  - $|\text{SE income}| > |\text{PE income}|$
- Compare empirical moments
  - PE median and mean incomes by age similar
  - SE median incomes by age similar
  - SE mean incomes by age starkly different

SE Median Income



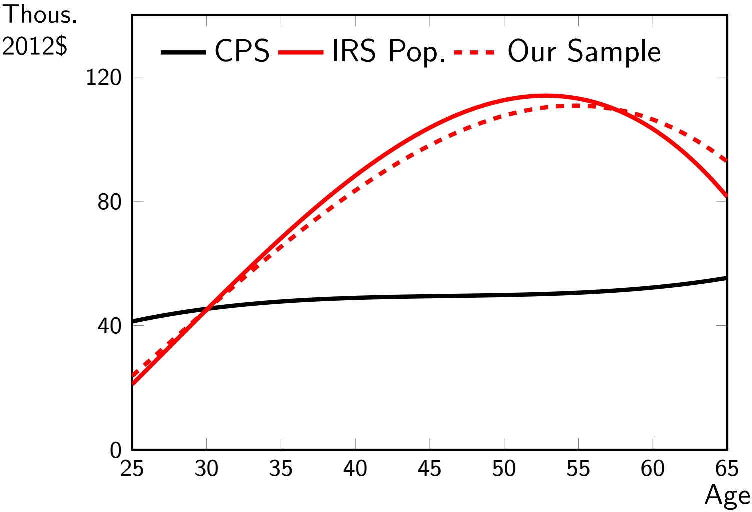
- Broadly similar

SE Mean Income



- Starkly different

SE Mean Income



- Starkly different, even for our balanced panel

Life-cycle Profile Estimation

Using longitudinal IRS data

Comparisons of Self- and Paid-Employed

- Central to the analysis is SE vs PE comparisons
- Idea:
  - Only self-employed rewarded for firm-specific investment
  - Can compare self- and paid-employed with
    - Same demographics, industry, education, etc.
    - Different investment opportunities
  - Look for differences in life-cycle income growth profiles

Object of Interest

$$\text{Income}(\text{Age} \mid \text{Individual and aggregate factors})$$

Challenges

- Selection
  - Incomes driven by latent characteristics
  - ⇒ Allow for unrestricted intercept
- Survival
  - Income higher because successful remain
  - ⇒ Study "attached" and "switchers" separately
- Identification
  - Time and age effects not separately identified
  - ⇒ Exploit overlapping cohorts
- Signs
  - Business incomes can be negative
  - ⇒ Estimate in levels with flexible error structure

Estimation Procedure

- Estimate time (β) and age (γ) effects 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

- Estimate time (β) and age (γ) effects 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 $\mathcal{I}$

- Requires assumptions to separately identify β and γ

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

A Practical Footnote: Easy to do

- Using least-squares approach

$$\min_{\{\Delta\beta_g,\overline{\gamma}_g^a\}} \sum_{g\in\mathcal{G}} \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{I}} \left(\Delta y_{it} - \Delta\beta_{g(i),t} - \overline{\gamma}_{g(i)}^{a(i,t)}\right)^2$$

$\Rightarrow$ Solving small linear systems for each g

$$\begin{pmatrix} \text{Population} \\ \text{Counts} \\ \text{for} \\ \text{different} \\ \text{ages and} \\ \text{times} \end{pmatrix} \begin{pmatrix} \Delta\beta_g^{2001} \\ \vdots \\ \Delta\beta_g^{2015} \\ \overline{\gamma}_g^{26} \\ \vdots \\ \overline{\gamma}_g^{65} \end{pmatrix} = \begin{pmatrix} \text{Avg.} \\ \text{Incomes} \\ \text{at} \\ \text{different} \\ \text{ages and} \\ \text{times} \end{pmatrix}$$

Application: set $\mathcal{G}$ with 35,117 subgroups

- Time-invariant characteristics include usual ones:
  - Cohort (50-59, 60-69, 70-75)
  - Gender (M/F)
  - Educated (yes/no)
  - Skilled cognitively, interpersonally, mannually (yes/no's)
  - Industry (20 2-digit)
  - Married (9 or more years, yes/no)
  - Children (have/don't have)
- Plus addition relevant for occupational choice:
  - *Employment attachment*

Employment Attachment

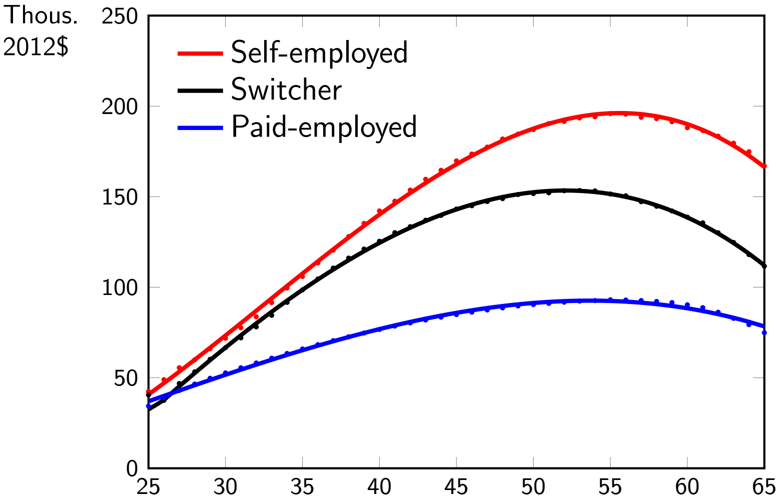
- *Attached* (SE or PE) if:
  - Same employment status for 12+ years
  - Fewer than 2 switches in status during sample
  - No intermediate spells of non-employment
- *Almost attached* (SE or PE) if:
  - Same employment status for 12+ years
  - More than 2 switches in status during sample
  - No intermediate 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

Employment Attachment

- Sample counts in millions
  - 36.1 attached to PE
  - 1.9 attached to SE
  - 0.3 almost attached to PE
  - 0.2 almost attached to SE
  - 3.2 mostly switchers
  - 23.3 any NE

Empirical Results:
Time and Age Effects

Income Profiles



- Does entrepreneurship pay? Yes

Understating the SE-PE Gap

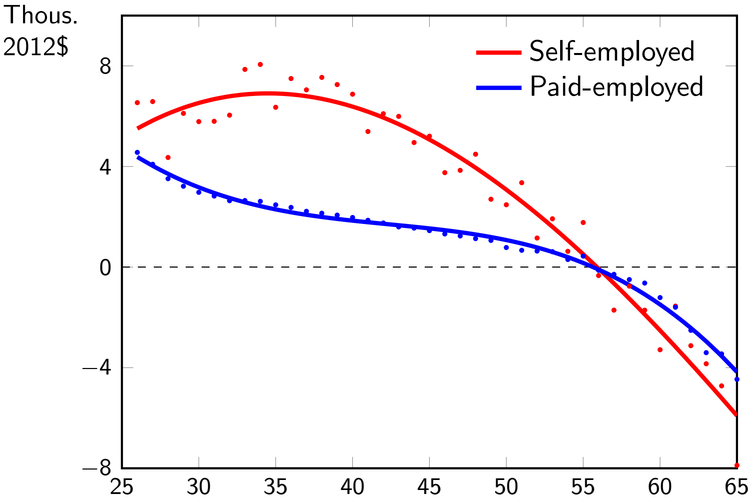
- SE-PE comparisons based on reported net incomes
- BEA (2012) estimates for aggregated pass-through income
  - Reported net income of 1,170 billion
  - Misreported net income of 698 billion

$\Rightarrow$ Scope for shrinking tax gap even from attached SE subgroup

Estimated Time Effects Relative to Total

- Flexible approach allows for differences in 2008-09

Estimated Growth for Attached SE and PE



- Significantly higher and more persistent growth for SE

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
    - Men

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services
  - Educated

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services
  - Educated
  - Interpersonally skilled

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services
  - Educated
  - Interpersonally skilled
  - Not manually skilled

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services
  - Educated
  - Interpersonally skilled
  - Not manually skilled
  - Not cognitively skilled

Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services
  - Educated
  - Interpersonally skilled
  - Not manually skilled
  - Not cognitively skilled
  - Attached to paid- or self-employment

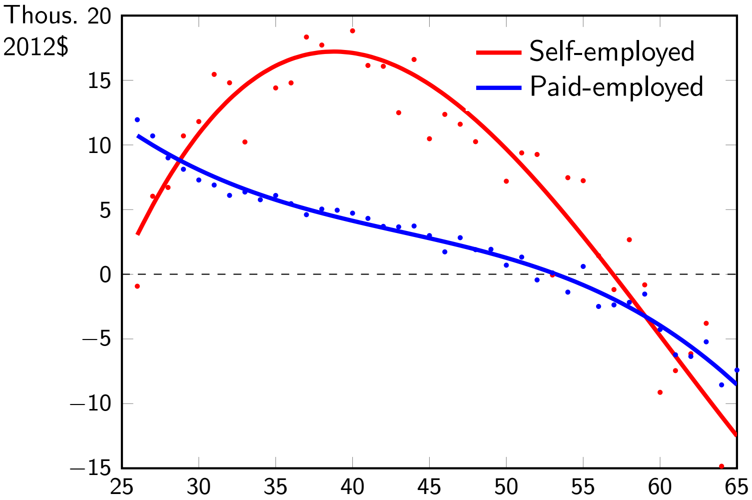
Disaggregating Trends

Rich data allows for disaggregated analysis

- For example:
  - Men
  - Married
  - Work in professional services
  - Educated
  - Interpersonally skilled
  - Not manually skilled
  - Not cognitively skilled
  - Attached to paid- or self-employment

Just two of our 35,117 groups

Estimated Growth For the Detailed Group



- Even more pronounced hump

Growth Gap Decomposition

Cumulative Share	Characteristics						
	Industry	Male	Married	Educated	Interpersonal	Cognitive	Manual
14.8							
26.9							
33.0							
39.0							
44.7							
49.9							
54.3							

- Small number of groups account for most of growth gap

Growth Gap Decomposition

Cumulative Share	Characteristics						
	Industry	Male	Married	Educated	Interpersonal	Cognitive	Manual
14.8	Health						
26.9	Prof.						
33.0	Health						
39.0	Finance						
44.7	Prof.						
49.9	Constr.						
54.3	Retail						

- Small number of groups account for most of growth gap

Growth Gap Decomposition

Cumulative Share	Characteristics						
	Industry	Male	Married	Educated	Interpersonal	Cognitive	Manual
14.8	Health	$\surd$	$\surd$				
26.9	Prof.	$\surd$	$\surd$				
33.0	Health	$\surd$	$\surd$				
39.0	Finance	$\surd$	$\surd$				
44.7	Prof.	$\surd$	$\surd$				
49.9	Constr.	$\surd$	$\surd$				
54.3	Retail	$\surd$	$\surd$				

- Small number of groups account for most of growth gap

Growth Gap Decomposition

Cumulative Share	Characteristics						
	Industry	Male	Married	Educated	Interpersonal	Cognitive	Manual
14.8	Health	√	√	√	√		
26.9	Prof.	√	√	√	√		
33.0	Health	√	√	√	√		
39.0	Finance	√	√	√	√		
44.7	Prof.	√	√	√	√		
49.9	Constr.	√	√	√	√		
54.3	Retail	√	√	√	√		

- Small number of groups account for most of growth gap

Growth Gap Decomposition

Cumulative Share	Characteristics						
	Industry	Male	Married	Educated	Interpersonal	Cognitive	Manual
14.8	Health	√	√	√	√	√	
26.9	Prof.	√	√	√	√		
33.0	Health	√	√	√	√	√	√
39.0	Finance	√	√	√	√		
44.7	Prof.	√	√	√	√	√	
49.9	Constr.	√	√	√	√	√	
54.3	Retail	√	√	√	√	√	

- Small number of groups account for most of growth gap

Empirical Results:
Tracking the Dollars

Tracking the Dollars

- For each industry, cohort, gender
  - Rank individuals by average income
  - Construct income shares by percentile
- Aggregate using population counts

Typical Dollar

Percentile Group	Income Share All	Self	Paid
$< 10^{th}$	0.8	-1.4	1.2
10^{th} to 25^{th}	4.5	3.0	4.7
25^{th} to 75^{th}	36.8	18.6	39.9
75^{th} to 90^{th}	21.8	15.8	22.8
$> 90^{th}$	36.1	64.1	31.4

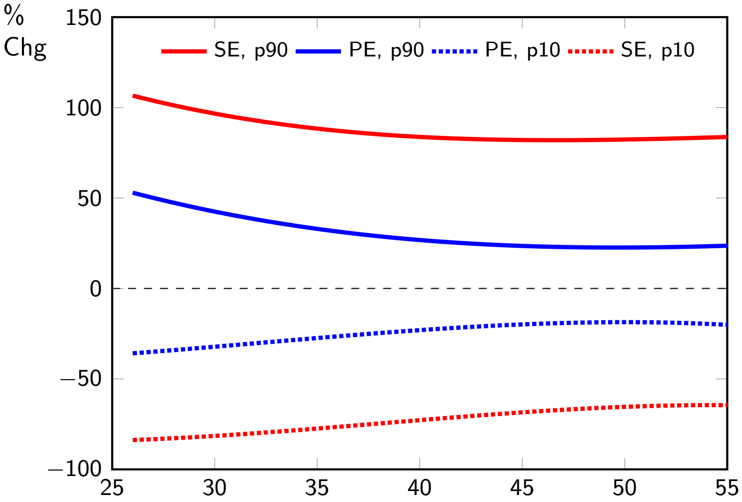
- 80% of entrepreneurial income
  - In 75+ percentile of income shares
  - Not well measured in survey samples

Empirical Results:
Volatility Patterns

Volatility Patterns

- Compute percentiles of $\Delta\epsilon_{i,a}/|y_{i,a-1}|$ after
  - Grouping all attached SE and all attached PE, **or**
  - Averaging subgroups of attached SE and attached PE
- Results show volatility
  - 2 to 3 times greater in SE than PE
  - Decreasing with age for both SE and PE
  - Almost all within group
- Next, plot income changes at 10^{th} and 90^{th} percentile

Income Changes for Attached SE and PE



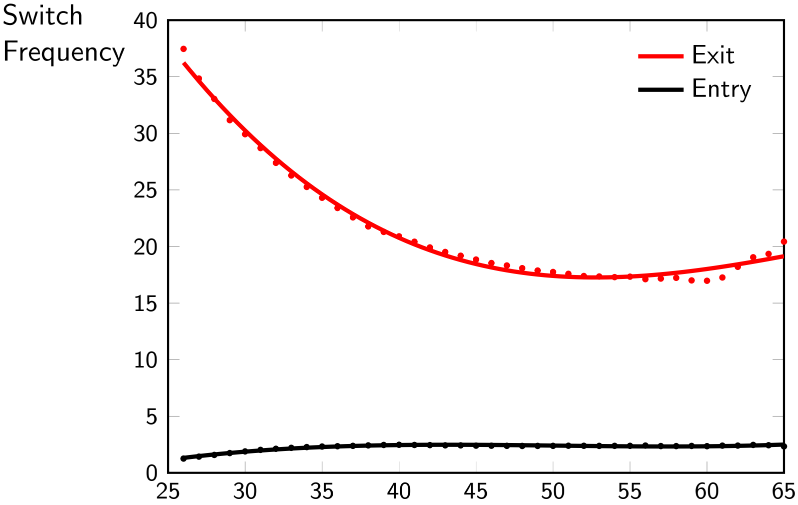
- Greater volatility in SE but decreasing with age

Entrepreneurial Choice: Entry and Exit

Entry and Exit

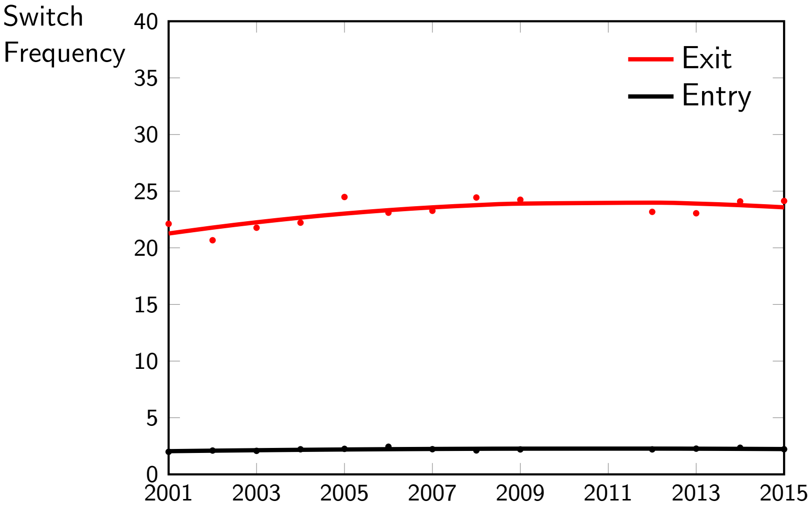
- Compute frequency of switches into/out of SE
  - By age
  - By year
- Results show
  - Similar magnitudes to survey estimates
  - Little noticeable change in Great Recession
- Next, plot results

Entry to and Exit from SE by Age



- Suggests early experimentation and learning

Entry to and Exit from SE by Year



- Suggests SE not a hedge against unemployment risk

Entrepreneurial Choice: Determinants of Self-Employment

Determinants of Self-Employment

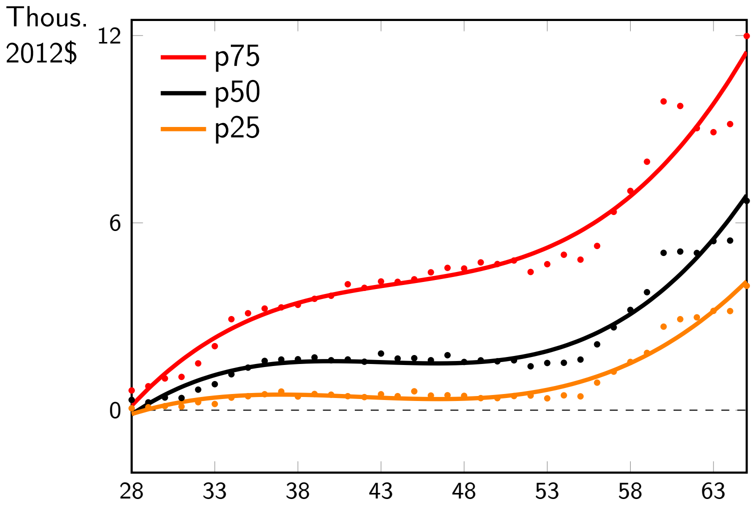
- Compare outcomes of SE entrants to "similar" peers
  - One-time entrants into SE ("Treatment")
  - Future switchers with same characteristics ("Control")
- Assess "misfit" hypothesis for SE
  - Have low past PE income
  - Use SE as fallback option

Determinants of Self-Employment

- Idea:
  - Compute average of x before switch (eg, $x =$PE income)
  - Compare x for i with that of matched peers $m(i)$
  - Use cohort, gender, NAICS for matches
- Compare differences Δ using 3-year past data:

$$\Delta_{it} = \frac{1}{3}\sum_{j} x_{i,t-j} - \frac{1}{3N_{m(i)}}\sum_{m(i)}\sum_{j} x_{m(i),t-j} \tag{1}$$

How Different are Past Wage Incomes?

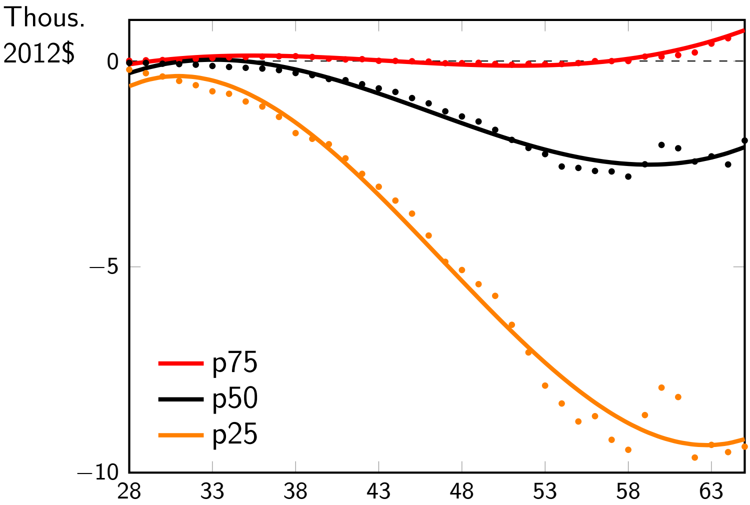


- Suggests wage income is *higher* for switchers before entry

Repeat Exercise with Asset Income

- Assess "financial-friction" hypothesis
  - Have high past income
  - Need financing to start businesses
- Condition also on percentile of past income

How Different are Past Asset Incomes?



- Suggests asset income is *lower* for switchers before entry

Informing Theory

Empirically-Motivated Features

- Two features suggested by empirical results:
  - Investment in self-created intangible assets
  - Incomplete information about entrepreneurial productivity
- Why self-created intangibles needed?
  - Owners invest time building customer-bases, brands, etc
  - Investment implies high, persistent income growth
- Why incomplete information needed?
  - Owners require time to learn their productivity
  - Learning implies declining exit rates

$\Rightarrow$ Added to decision theoretic problem dynamic program

A Theoretical Case Study: Young Entrepreneurs

Predictions for Young Entrepreneurs

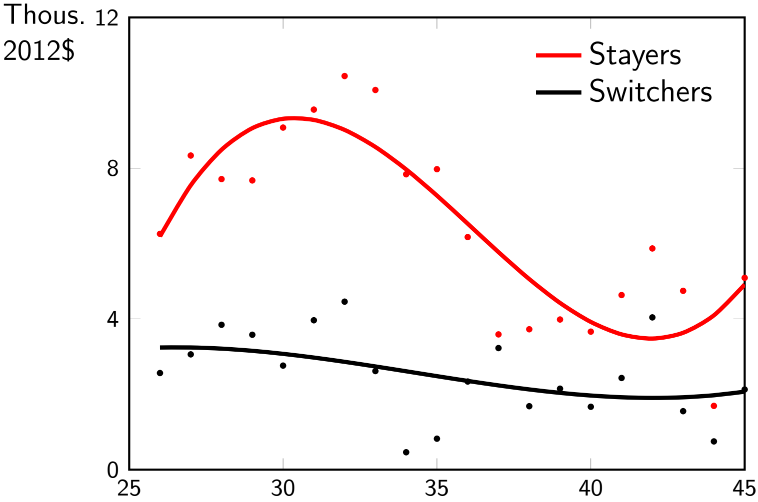
- Choose parameters consistent with IRS micro data
- Simulate model time series over the life cycle
- Aggregate simulations using IRS counts and entry ages
- Construct growth profiles for young SE stayers/switchers

Predictions for Young Entrepreneurs

- Choose parameters consistent with IRS micro data
- Simulate model time series over the life cycle
- Aggregate simulations using IRS counts and entry ages
- Construct growth profiles for young SE stayers/switchers

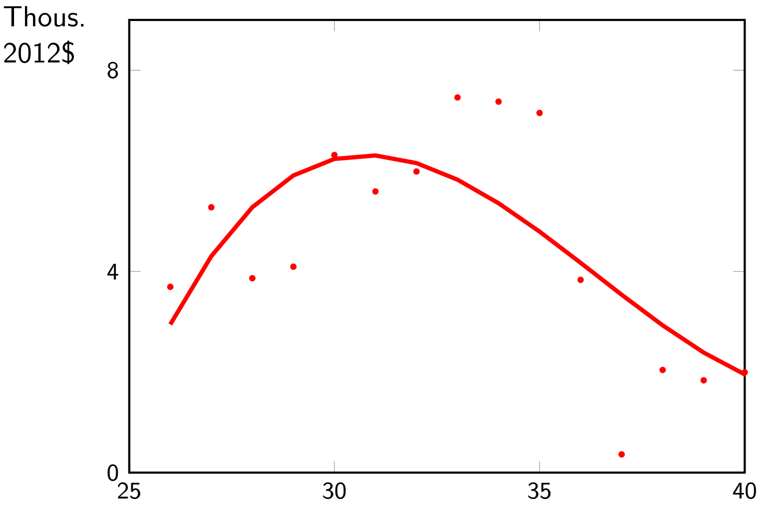
Let's start with the data...

1970-75 Cohort with 5+ Years SE Experience



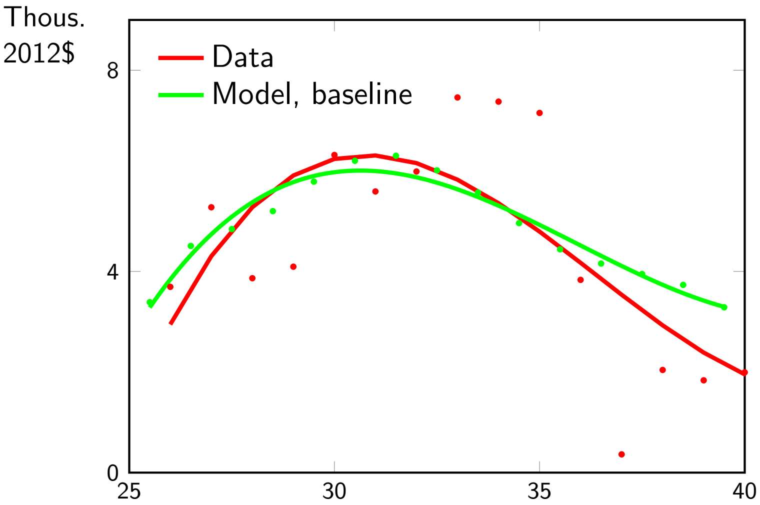
- Use results to construct growth differential for data

Growth Differential for Young Entrepreneurs



- Suggests SE stayers different than switchers even early on

Growth Differentials for Young Entrepreneurs



- Theory generates comparable growth pattern

Summary

- 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

Dynamic Program

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

where

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

$j' = j + 1$ and j is age

$\epsilon' =$ a Markov chain given productivity

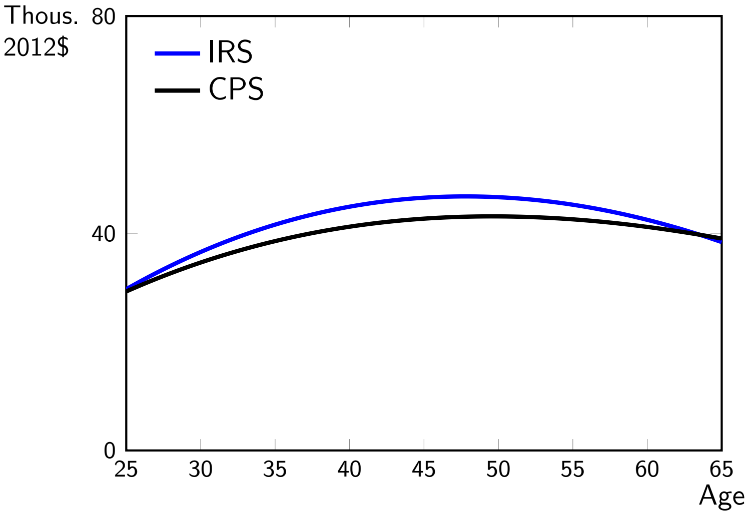
$z_j = \bar{z}_0 + \eta_j$ given $\eta_j \sim N(0, \sigma_\eta^2)$ and productivity z_j

$\mu_j = \mu_{j-1} + \sigma_{j-1}^2 (z_{j-1} - \mu_{j-1})/(\sigma_{j-1}^2 + \sigma_\eta^2)$

$\sigma_j^2 = \sigma_{j-1}^2 \sigma_\eta^2 / (\sigma_{j-1}^2 + \sigma_\eta^2)$

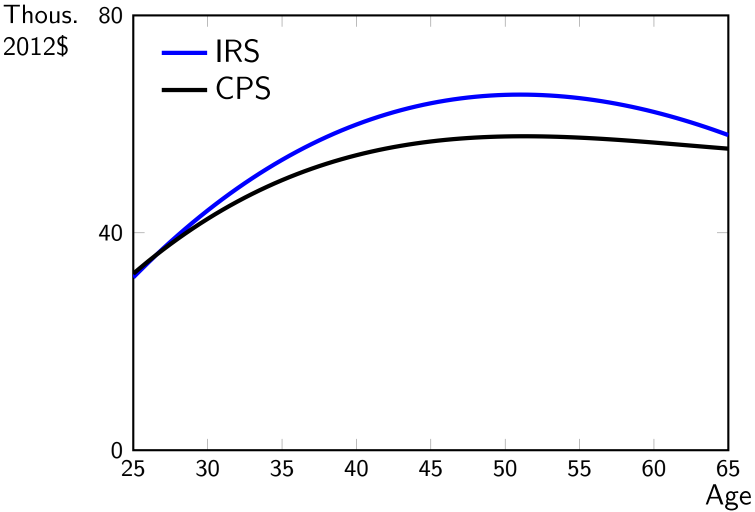
back

PE Median Income



Back

PE Mean Income



Back