On the Nature of Entrepreneurship

A. Bhandari,\textsuperscript{1} T. Kass,\textsuperscript{2} T. May,\textsuperscript{1} E. McGrattan,\textsuperscript{1} and E. Schulz\textsuperscript{3}

\textsuperscript{1}Department of Economics
University of Minnesota

\textsuperscript{2}Office of Tax Analysis
Department of Treasury

\textsuperscript{3}Internal Revenue Service
Department of Treasury
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
Data
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:**
  - $|\text{SE income}| > 5,000$ in 2012$ \text{ and at least one of:}$
    - $|\text{SE income}| > \text{PE income}$ or
    - Share of gross profits $> \text{PE income}$ or
    - Share in business $\times$ employees $\geq 1$

- **Paid-employed (PE) in a given year if:**
  - Not SE
  - $\text{PE income} > 5,000$ 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

\[ \text{Income(} \text{Age} \mid \text{Individual and aggregate factors}) \]
Estimation Procedure

- Statistical model for income:

\[ y_{it} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a(i,t)} \gamma_{c(i),g(i)}^a + \epsilon_{i,t} \]

where

- \( i \in I \) is set of individuals
- \( t \in T \) is set of calendar dates
- \( c \in C \) is set of birth years
- \( a \in A \) is set of ages
- \( g \in G \) is set of groups partitioning \( I \)
Estimation Procedure

- **Statistical model for income:**

\[ y_{it} = \alpha_i + \beta g(i),t + \sum_{a=a_0}^{a(i,t)} \gamma_{c(i),g(i)}^a + \epsilon_{i,t} \]

where

- \( i \in I \) is set of individuals
- \( t \in T \) is set of calendar dates
- \( c \in C \) is set of birth years
- \( a \in A \) is set of ages
- \( g \in G \) is set of groups partitioning \( I \)
Estimation Procedure

- Statistical model for income:

\[ y_{it} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a(i,t)} \gamma_{c(i),g(i)}^a + \epsilon_{i,t} \]

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:

\[ y_{it} = \alpha_i + \beta g(i),t + \sum_{a=a_0}^{a(i,t)} \gamma_{c(i),g(i)}^a + \epsilon_{i,t} \]

where

- \( i \in I \) is set of individuals
- \( t \in T \) is set of calendar dates
- \( c \in C \) is set of birth years
- \( a \in A \) is set of ages
- \( g \in G \) is set of groups partitioning \( I \)
Estimation Procedure

• Estimation of time ($\Delta \beta$), age ($\gamma$) effects:

$$\Delta y_{i,t} = \Delta \beta g(i,t) + \gamma a(i,t) + \gamma c(i),g(i) + \Delta \epsilon_{i,t}.$$  

• Identification:

  ○ Assume that age effects are constant across binned cohorts

  ○ Normalize time effects to reflect group-specific growth
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
Comparisons Central for Analysis

● Attached SE vs PE growth informs differences in:
  ○ Preferences for amenities, risk, etc.
  ○ Productivities in SE/PE
  ○ Investment opportunities
  ○ Non-compliance opportunities

● Switchers vs non-switchers informs entrepreneurial choice
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

⇒ Entrepreneurial investment does pay
Income Profiles: Attached Subsamples

Graph showing the income profiles for Self-employed and Paid-employed across different ages in thousands of 2012 dollars.

- Self-employed: Red line
- Paid-employed: Blue line

Axes:
- Y-axis: Thousands of 2012$ (0 to 250)
- X-axis: Age (25 to 65)
Growth Profiles: Attached Subsamples

Thousands of 2012$

Age

Self-employed
Paid-employed

Time Effects
Disaggregating the Main Results for Attached Subsamples
Growth Profiles: By Gender

A. Men

B. Women

Thousands of 2012$

Age

Self-employed
Paid-employed
Growth Profiles: By Education

A. College-educated

B. Not

Self-employed
Paid-employed
Growth Profiles: By Industry

A. Professional Services

B. Health Care

Thousands of 2012$

Age

SE

PE
Growth Profiles: By Income Ranks

A. Top 25%

B. Bottom 25%

Thousands of 2012$

Age

Self-employed
Paid-employed

25 30 35 40 45 50 55 60 65

25 30 35 40 45 50 55 60 65
Down to the Subgroup Level: An Example

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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: 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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: An Example

- Consider
  - Men
Down to the Subgroup Level: An Example

- Consider
  - Men
  - Married
Down to the Subgroup Level: 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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
Down to the Subgroup Level: An Example

- Consider
  - Men
  - Married
  - With kids
  - Educated
  - Not cognitively skilled
Down to the Subgroup Level: 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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: An Example

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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: An Example

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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: 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

⇒ Just 2 of the 46,080 groups
Down to the Subgroup Level: 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

⇒ Just 2 of the 46,080 groups
Growth Profiles: Example subgroups

- Self-employed
- Paid-employed

Thousands of 2012$

Age
Volatility Patterns

- Large literature on risk in entrepreneurship
  - 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\textsuperscript{th} and 90\textsuperscript{th} percentiles by age and employment status
Income More Volatile for Attached SE

SE ≈ 3 times more volatile
Income More Volatile for Attached SE

Age vs. Income Change (%)

- SE p90
- SE p10
- PE p90
- PE p10

Volatility declining with age
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 $\lambda$ 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 $\rho \to 0$
    \[
    V_t(\{c_j\}_{j=t}^{\infty}) = [(1 - \beta)c_t^\rho + \beta(E_t V_{t+1}^\alpha)^{\rho/\alpha}]^{\frac{1}{\rho}}
    \]
  - 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 $\lambda$ sacrificed to fully insure $c = y$
  \[
  \lambda = -0.5\alpha\beta\sigma_r^2
  \] (≈ $Q^2 = 9$)
Analysis of Entrepreneurial Choice with Full Sample
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

![Graph showing past wage incomes for switchers]
Past Asset Incomes Lower for Switchers

The graph shows the past asset incomes over age for different percentiles. The red line represents the 75th percentile, the black line the 50th percentile, and the blue line the 25th percentile. The incomes are measured in thousands of 2012$. The age is presented in years, ranging from 28 to 63.
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
Relation to Survey-based Findings
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
  - 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 using SE as fallback
  - Face few liquidity constraints
Problems Even with Larger Surveys

- Compare cross-sectional moments:
  - IRS full population
  - Current population household survey

- Use comparable employment-status categorization:
  - SE if $|\text{SE income}| > 5,000$ and $|\text{SE income}| > \text{PE income}$
  - PE if not SE and PE income $> 5,000$ in 2012$
  - NE if not SE or PE
Empirical Moments: IRS vs CPS (Th. 2012$)

SE Median Income

PE Median Income

SE Mean Income

PE Mean Income
Informing Theory: Details
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
  - 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

\[
V_k(s) = \max \{ U(c, \ell) + \beta EV(s') \}
\]

\[
a' = (1 + r)a + pez f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0
\]

\[
\kappa' = (1 - \delta_\kappa)\kappa + f_\kappa(h_\kappa, e)
\]

\[
\ell = 1 - h_y - h_\kappa
\]

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

- 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

\[
V_k(s) = \max \{ U(c, \ell) + \beta EV(s') \}
\]

\[
a' = (1 + r)a + pe^z f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0
\]

\[
\kappa' = (1 - \delta_\kappa)\kappa + f_\kappa(h_\kappa, e)
\]

\[
\ell = 1 - h_y - h_\kappa
\]

• Two production technologies:
  
  o \( f_y(\kappa, h_y, k, n) \): goods and services
  
  o \( f_\kappa(h_\kappa, e) \): new intangible assets
Modeling Intangibles

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

- Dynamic program for entrepreneur

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

- 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**

\[
V_k(s) = \max \{U(c, \ell) + \beta EV(s')\}
\]

\[
a' = (1 + r)a + pe^z f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn - e - c \geq 0
\]

\[
\kappa' = (1 - \delta_{\kappa})\kappa + f_{\kappa}(h_{\kappa}, e)
\]

\[
\ell = 1 - h_y - h_{\kappa}
\]

- **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

Thousands of 2012$ Data

- Black line: Data
- Red line: Model

Axes:
- X-axis: 25 to 40
- Y-axis: 0 to 8

Legend:
- Data
- Model
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 ($\geq 2$)
  - Average time effect satisfies (where $\bar{y}_{g,t_0}$ is avg income for $g$):
    \[
    \frac{\Delta \beta_g}{\bar{y}_{g,t_0}} = \frac{\mu_g}{T} \sum_t (1 + \mu_g)^t
    \]
- Allows flexibility when set $G$ large
Employment Attachment

- **Attached (SE or PE) if:**
  - Fewer 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
Income Profiles: Add Switchers

- Self-employed
- Paid-employed
- Switchers

Income in Thousands of 2012$: Age 25 to 65

- Self-employed: Red line
- Paid-employed: Blue line
- Switchers: Black line

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
Time Effects Relative to Income

self-employed
paid-employed

Percent vs. Tax Year