On the Nature of Entrepreneurship*

Anmol Bhandari
University of Minnesota
bhandari@umn.edu

Tobey Kass
Office of Tax Analysis
tobey.kass@treasury.gov

Thomas J. May
University of Minnesota
may00013@umn.edu

Ellen R. McGrattan
University of Minnesota
erm@umn.edu

Evan Schulz
Internal Revenue Service
evan.d.schulz@irs.gov

July 12, 2023

ABSTRACT

This paper elucidates the nature of entrepreneurship by studying the life-cycle income profiles and entry and exit behavior of self-employed individuals. Results are based on U.S. administrative data from the Internal Revenue Service and Social Security Administration over the period 2000–2015 for subgroups of the population differing by gender, marital status, education, occupation, industry, cohort, and employment status. Contrary to top-coded survey evidence based on relatively small samples and short panels, we find that self-employed individuals have significantly higher average income and steeper, more persistent income growth profiles than their paid-employed peers with similar characteristics. Contrary to survey evidence, we find that new entrants into self-employment have higher labor incomes and lower asset incomes prior to entry relative to similar peers that do not enter. We develop a theory of entrepreneurial choice and compare it to the subsample of young entrepreneurs in our data. We find that including firm-specific investment, risky incomes, and incomplete information is necessary for the theory to match the observed income growth profiles and switching behavior.

* Corresponding author: McGrattan, University of Minnesota, 1925 4th Street South, Minneapolis, MN 55455. 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. We acknowledge support from the National Science Foundation (Award #2214248). For their valuable feedback on earlier drafts, we thank discussants Peter Klenow and Ross Levine and seminar participants at the Office of Tax Analysis, Federal Reserves, Stanford, Harvard, Yale, Princeton, ASU, NYU, and Toronto and conference participants at NBER, SED, and Hoover Institution. 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). This research was conducted while Kass was an employee at the U.S. Department of the Treasury. Any findings, interpretations, opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views or political positions of the IRS or the U.S. Department of the Treasury, or the NSF. 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.
1 Introduction

Despite volumes written on the topic, there is surprisingly little concordant evidence about returns to entrepreneurship.\(^1\) The goal of this paper is to fill in part of the gap in our knowledge using U.S. administrative tax filings of employees and business owners over the period 2000–2015. We develop an econometric framework to estimate growth in incomes and use it to compare the life-cycle income profiles and switching behavior of individuals who share similar characteristics but differ in their choice of self- or paid-employment. We then use these statistics to inform economic theories of occupational choice and entrepreneurship.

We construct life-cycle income profiles for groups of individuals with different demographic and labor market characteristics. We utilize the Statistics of Income (SOI) Databank, which combines records from the Social Security Administration (SSA) and the Internal Revenue Service (IRS), providing us with demographic information such as age, gender, marital status, and the number of children, as well as information on employment status, occupation, industry, own incomes, and household incomes. We use machine learning algorithms to impute additional information such as educational attainment and occupation to measure a broad notion of skills. Our measure of self-employment income is the sum of incomes from proprietorships (Form 1040, Schedule C net profits), partnerships (Form 1065, Schedule K-1 ordinary business income), S corporations (Form 1120-S, Schedule K-1 ordinary business income), and own-business compensation (Form W-2 wages). Paid-employment income is wage income (Form W-2) less any own-business compensation. Along with the income measures, we use auxiliary data such as number of employees and other business deductions to classify individual-year pairs as self-employed, paid-employed, or non-employed.

Our first exercise is to estimate a rich specification of income over the life cycle that uses information on the group and cohort of individuals. It has three components. The first component is an individual-level fixed effect meant to capture latent abilities, preferences, and other unobservable characteristics. The second component is a time effect that depends on the individual’s cohort and group and is meant to capture changes in income specific to our sample, such as the large recession occurring in 2008–2009. The third component is an age effect that depends on the individual’s cohort and group and is meant to capture changes in income over the life cycle as individuals gain more experience on the job. Our identification scheme assumes age effects are similar across binned cohorts and, with differenced income from our dataset, allows us to estimate the time and age effects for all subgroups.

To overcome issues related to compositional bias, we separately study individuals who switch frequently between self-, paid-, and non-employment and those who are more attached in their employment status. By “attached,” we mean that the individual has the same employment status

\(^1\)See Parker (2018) for a comprehensive review of the literature.
for 12 or more years, with at most two switches in status over the 16-year sample period and no intermediate spells of non-employment. Contrasting the income profiles of those who are attached with those who are less attached allows us to discern pecuniary versus non-pecuniary motives, both of which underlie an individual’s choice of being attached to an activity.

For the full sample of attached employees and entrepreneurs, we find that growth in employee income declines across the life cycle, whereas growth in entrepreneurial income remains persistently high until mid-career and then gradually declines. At age 25, the mean incomes are $34 thousand (in 2012 U.S. dollars) for the paid-employed and $42 thousand for the self-employed. By age 55, the self-employed are earning more than twice the paid-employed, roughly $210 thousand versus $89 thousand. If we decompose the aggregate differences in income growth between the self- and paid-employed in their 30s and attribute them to subgroups of our sample, we find the key contributors driving these differences are married males with occupations requiring education and interpersonal skills that have jobs in health care, professional services, finance, retail trade, and construction. For these subgroups, the growth is highest during the mid-30s, suggesting that business owners make initial investments to build a business (as in Bhandari and McGrattan (2021)) or experiment early in their careers in order to learn their productive capabilities in different occupations (as in Jovanovic (1982)). Investment and experimentation would delay growth and generate the delayed hump-shaped growth profiles that we observe. We also find that the profiles of individuals who have non-employment spells and are relatively attached to non-market work—a large fraction of our sample—are flatter and substantially lower than those attached to paid-employment or self-employment.

Our findings on life-cycle growth rates for self-employed individuals are different from those in the survey-based literature (see Hamilton (2000) and Hurst and Pugsley (2011)), who find flatter profiles and conclude that self-employment has a large, non-pecuniary role. We reconcile those differences by comparing distributions of income by age and employment across the Current Population Survey with the IRS administrative data. We find that for comparable self-employed individuals, the average based on IRS data is as large as $42 thousand more than the estimate based on CPS data. In contrast, the paid-employed averages are barely different across datasets for prime-age individuals. Furthermore, the CPS-IRS differences in median incomes by age and employment status are much smaller. We conclude that surveys fail to capture well the right tail of the income distribution, which for the self-employed contains most of the income. Thus, while the survey-based analysis paints a reliable picture of the median self-employed individual, it is not reflective of how the median dollar in self-employment is earned.

After analyzing the mean growth rates, we study the variability and persistence of income changes to investigate the risky nature of entrepreneurship. We focus on the dispersion and autocorrelation of income changes—two statistics that have been used in previous work to quantify
the gains of greater insurance against idiosyncratic risk. If we measure dispersion using the 90–10 difference in percentage growth (after netting out time and age effects), we find that the volatility is 3 times greater in self-employment relative to paid-employment, whereas the autocorrelations of the rates are roughly the same across employment status. While greater variability translates into higher welfare gains for smoothing entrepreneurial risk, we find that individuals in our sample have means to smooth consumption expenditures through spousal wages and other household income.

To gain a better understanding of entrepreneurial choice, we analyze entry into and out of self-employment over the life cycle. For our sample, exit rates decline significantly over the life cycle, with experimentation in entrepreneurship occurring at younger ages, but are flat across time. Entry rates are flat over both the life cycle and across time. Remarkably, we see little change during the 2008–2009 recession, which suggests that entrepreneurship was not used as a fallback option. Relatedly, if we compare past labor incomes for observationally similar individuals—one entering self-employment and the other not—we find the newly self-employed had higher past income, which is inconsistent with the view that “misfits” are pushed into entrepreneurship. If we instead compare past asset incomes for these observationally similar individuals, we find the opposite: the newly self-employed had lower past asset income, which is inconsistent with the view that entrepreneurs face liquidity constraints (see Evans and Leighton (1989)). If we compare the earnings of those that switch employment status—whether they are switching from self- to paid-employment or vice versa—with those observationally similar peers that do not switch, we find nearly as many increases as decreases in income, suggesting that both pecuniary and non-pecuniary motives drive occupational choice.

Having documented the key empirical patterns of our sample, we use theoretical predictions of an occupational choice model to interpret our findings. In the model, our theoretical entrepreneurs spend some time investing in self-created intangible assets—for example, customer bases and trade names—and growing to an optimal size. Self-employment carries risks, and young entrepreneurs start with little to no financial assets or other incomes that can be used to smooth consumption during the first years. Meanwhile, productive abilities must be learned, and when they are, exit due to selection occurs. If exit does occur, the business is sold, intangible assets are transferred, and the owner switches to paid-employment.

Because we are interested in the role of investment and experimentation in generating realistic growth profiles and hazard rates corresponding to entrepreneurship, we compare model simulations to the youngest cohort of our sample—those born between 1970 and 1975—that are self-employed for at least five consecutive years prior to age 35. For the simulations, we use the baseline parameterization of Bhandari and McGrattan (2021), who abstract from learning, and then use moments from the IRS subsample to set parameters of the learning process and income shocks. This parameterized version of the model is shown to generate profiles consistent with young entrepreneurs in
our IRS sample. We find that learning is a necessary feature of the model: if there is too much certainty about business owners’ productive capabilities and the nature of business risk, then occupational choices are made quickly. In that case, the model cannot rationalize self-employment stints as long as five years followed by a switch. Similarly, we find that firm-specific investment is a necessary feature: if an owner only requires factor inputs that can be rented or hired without delay, then the business can immediately be scaled to its optimal size. In this case, the model cannot rationalize persistent differences in income growth when comparing the profiles for entrepreneurs that continue in business and those that exit.

An important by-product of our work is a longitudinal database of business owners that can be used to develop predictive tools—both theoretical and statistical—for improved tax administration. This database allows for a broader scope of analysis, beyond what is possible with survey data alone. With surveys, researchers can study the typical entrepreneur, whereas we can study the typical dollar earned in self-employment and can track the individual earning it over a long period. What our analysis shows is that the typical dollar is earned by those with incomes in the top 25 percent that are attached to self-employment, and these individuals have life-cycle income profiles that are easily distinguishable from their paid-employed peers.

2 Data

In this section, we describe our main sample drawn from U.S. administrative tax records.\textsuperscript{2} We start with details of the data source and definitions of self- and paid-employment income. We then describe algorithms to impute skill and education levels.

2.1 Sample

When constructing our sample, we start with records in the SOI Databank, which is a de-identified balanced panel of all living individuals with a U.S. Social Security number over the period 1996 to 2015.\textsuperscript{3} For each individual there are rows—one for each year—and columns recording demographic information from the SSA (such as age and gender) and economic data from tax filings (such as information on individual income tax forms and attachments). This database is our primary source for data.

The SOI Databank includes information on wages and salaries reported to the IRS on Form W-2 for employees and household-level Schedule C income reported on Form 1040. For individual proprietors, we assign incomes separately by Social Security number. For owners with pass-through businesses—partnerships and S corporations—we merge in information from Schedule K-1 filings.

\textsuperscript{2}Replication codes and detailed documentation are available at the IRS.

\textsuperscript{3}See Chetty et al. (2018) for full details on this database. We remove any person from our sample who died prior to 2015.
attached to Form 1065 and 1120-S, respectively. The Schedule K-1 data are available beginning in 2000, and thus our sample period ranges from 2000 to 2015. Because self-employment income must be reported on the standard Form 1040 when filing individual income taxes, we exclude from our baseline sample any individuals that exclusively use the simpler Forms 1040A or 1040EZ.

To construct income profiles by age, we use records for all individuals between the ages of 25 and 65 in the SOI Databank for the years 2000 through 2015—namely, birth cohorts 1950 through 1975. This balanced panel includes roughly 128 million individuals for 16 years (that is, 2 billion person-year observations). Another restriction we place on the sample is the availability of occupational information, which is used to impute levels of education and skill that play an important role in income determination. This restriction narrows our sample to roughly 80 million individuals. Details of the imputations are provided in Section 2.3.

### 2.2 Income Measures

For each individual-year observation, we compute two sources of income. The first is a measure of *self-employment income* and is defined as the sum of net profit or loss of sole proprietors (Form 1040, Schedule C, Line 31), the individual’s share of ordinary business income from partnerships (Form 1065, Schedule K-1, Part III, Line 1), the individual’s share of ordinary business income from S corporations (Form 1120-S, Schedule K-1, Part III, Line 1), and finally the individual’s income paid by the S corporations that they own as wages (Form W-2, Box 1). The second is a measure of *paid-employment income* and defined as the wages and salaries (on Form W-2, Box 1) paid by businesses that are not owned by the wage earner. We refer to the sum of self- and paid-employment income as *total income*, although it does not include other categories of adjusted gross income on the tax forms. These measures are computed before tax and transfers, exclude most employer fringe benefits, and are deflated by the Bureau of Economic Analysis’s (BEA) personal consumption expenditure price index and reported in thousands of 2012 U.S. dollars. No adjustments are made to account for potential income underreporting.

Although individuals can have both paid-employment income and self-employment income, we assign individuals to distinct employment categories each year based on a test designed to gauge their primary activity. To do that, we construct three categories: self-employed (SE), paid-employed (PE), or non-employed (NE) using the following definitions.

---

4Business net incomes of Subchapter C Corporate shareholders are not passed through to individual income tax forms until the companies distribute dividends or capital gains. Until then, there are no administrative records that can be attributed to the individual owners.

5The full SOI Databank sample over 2000–2015 has 7 billion observations and 3 billion for ages 25 to 65 if we include individuals that are not in our sample for all years.

6Here, we omit capital gains as a source of self-employment income, although we acknowledge that there could be realized gains from sales of intangible assets reflecting entrepreneurial investment that should be included with self-employment income. Including such gains only strengthens our main finding.
Definition 1. An individual-year pair is classified as self-employed (SE) if the absolute value of self-employment income exceeds $5,000 (in 2012 dollars) and any of the following conditions is true: (i) the absolute value of their self-employment income is greater than their paid-employment income; (ii) the sum across businesses of the individual’s ownership share times the number of its employees is larger than 1; or (iii) the sum across businesses of the individual’s ownership share of gross profits (receipts less cost of goods sold) is in excess of the individual’s paid-employment income.

We take the absolute value of the income because young entrepreneurs incur significant expenses when building up their businesses, and many have losses. The second additional criterion is added because hiring employees is indicative of owner attachment to self-employment. The third criterion allows for the fact that many successful business owners pay themselves little income to minimize taxes but earn incomes later when selling their businesses.

Our notion of self-employment is distinct from papers such as Smith et al. (2019), DeBacker, Panousi, and Ramnath (2022), Garin, Jackson, and Koustas (2022), and Lim et al. (2019), who all use IRS data to study business incomes. Smith et al. (2019) classify all individual recipients of K-1 as self-employed. Our definition excludes 43 million of the 138 million individual-year K-1 recipients in our sample from being classified as self-employed. These are cases in which an individual probably spends very few hours running a business and receives very little income from business filings. While this is not a concern for top incomes, which is the focus of Smith et al. (2019), our focus is to learn about returns to entrepreneurship. Therefore, we deliberately use a more conservative test when categorizing entrepreneurial activity. DeBacker, Panousi, and Ramnath (2022) use a panel that tracks tax filers for up to 32 years using the SOI sample from 1987. While this has the benefit of being a long panel, the number of self-employed individuals that are studied shrinks down to about 2,000 observations over a few cohorts. Such a restrictive sample would be unsuitable for achieving our two main goals: (i) to calculate life-cycle income profiles using overlapping cohorts to infer time and age effects and (ii) to understand the determinants of self-employment by comparing outcomes for narrowly defined groups—some of whom enter self-employment and some of whom do not. Garin, Jackson, and Koustas (2022) focus on Schedule SE filers. This focus is not suitable for our analysis because it misses entrepreneurs who make losses and S corporation owners that do not file Schedule SE—a significant fraction of business owners. Lim et al. (2019) focus on independent contractors that receive a Form 1099 and have less than $10,000 in deductions, excluding car and travel expenses. While these individuals are included in our sample, restricting the analysis to this group would eliminate a significant fraction of self-employment income. For example, for sole proprietors in our sample that satisfy the definition of independent contractor and have greater
than $5,000 in total receipts, the share of receipts from Form 1099 is only 6 percent.\(^7\)

Next we define paid- and non-employed categories.

**Definition 2.** An individual-year pair is categorized as *paid-employed* (PE) if it is not already categorized as self-employed and if the paid-employment income of the individual in that year exceeds $5,000 (in 2012 dollars).

**Definition 3.** An individual-year pair is categorized as *non-employed* (NE) if it is not already categorized as SE or PE.

To distinguish observations that are non-employed from those that are actually paid- or self-employed but missing in the SOI Databank, we use auxiliary data from Form 1040 and the individual’s Form W-2, Schedule C, or Schedule K-1 if any of these filings are available. Consider wage earners first. We compute a *wage gap* as the difference in wages and salaries reported on Form 1040 and the aggregated Box 1 wages on Form W-2 for the individual and spouse. If this gap is less than $1,000, we use the Form W-2 data for the individual. If the wage gap is greater than $1,000 and the individual is not married filing jointly, then we use the Form 1040 wage. If the wage gap is greater than $1,000 and the individual is married filing jointly, then we need to consider two cases. First, there may be Form W-2 filings for the spouse but not the individual. In this case, we use the Form 1040 wage less the spouse’s total W-2 wages. Second, the Form W-2 filings may be missing for both the individual and spouse. In this case, we take pro-rata shares for the two spouses based on tax filings in other years showing a wage gap of less than $1,000. If such information is unavailable, we split the Form 1040 wages evenly for the two spouses. In the case of business owners, we follow the same procedures as for wage earners, but in this case we use the owner’s Schedule C or Schedule K-1.

In the first column of Table 1, we provide summary statistics for our main sample. The full sample has 80 million individuals. Summing across all individual-year observations, we find the shares of self-, paid-, and non-employed are 8 percent, 74 percent, and 18 percent, respectively. They earn an average of $49 thousand (in 2012 dollars) in combined paid- plus self-employment income, with a range across the distribution from $6 thousand at the 10\(^{th}\) percentile to $92 thousand at the 90\(^{th}\) percentile—roughly a factor of 15. If we sum up self- and paid-employment incomes across all individual-year observations in our sample, we find the share of self-employment income is equal to 13 percent.\(^8\) The mean paid-employed income lies between 50th and 75th percentiles of the distribution of PE income, while the mean self-employed income is closer to the 90th percentile.

\(^7\)These results could change as the sample is expanded to include more recent tax years when there was significant growth in online platform employment.

\(^8\)If we include all earnings when computing this share, we find that self-employed individuals earn 16 percent of the total income.
of SE income, indicating a substantial right-skewness of the self-employment income distribution for the whole population.

2.3 Imputations for Skills and Education

A large empirical labor literature focuses on skills and education as determinants of income. In this section, we impute information on education and occupation, which is not readily available in the tax data for all individuals. We later use these estimates when analyzing subgroups of the population of tax filers.

2.3.1 Skills

After signing and dating the tax form, individual tax filers and their spouses are asked to self-report their occupation, which is summarized in the IRS data as a character string. The occupation information is available for e-filed returns for tax year 2005 and later, with the exception of 2012. For the sample of individuals born between 1950 and 1975, 89 million individuals e-filed at least once in the years these occupation strings are available. We are able to assign skill values to the subset of 80 million individuals in our main sample.

First, 73 million individuals provide usable occupations, which can be mapped directly to a standard occupational classification (SOC) code. For these individuals, we assign skill values using the procedure of Lise and Postel-Vinay (2020). The idea is to create a mapping between the SOC codes assigned to individuals and their cognitive, interpersonal, and manual abilities. This is done with the aid of the Occupational Information Network (O*NET) summary of skill requirements needed for each occupation. Since the summary of requirements is long for each occupation, Lise and Postel-Vinay (2020) use a principal component analysis (PCA) to construct indices—keeping the top three (orthonormal) components and ensuring that occupations requiring mathematics are encoded as “cognitive” skills, occupations requiring social perceptiveness are encoded as “interpersonal” skills, and occupations requiring mechanical knowledge are encoded as “manual” skills.

Second, there are 7 million individuals for whom we impute a skill value. For these individuals, we apply a $k$-nearest neighbor classifier for the imputation using information on $k$ “neighbors” from the subsample of the 73 million individuals that have a valid SOC code and assigned skill values. The neighbors share the same gender, marital status, birth cohort, and two-digit NAICS industry code and are nearest in the paths of employment status and incomes. For each subgroup, we operationalize choosing near neighbors in the case of time-varying income variables by applying a PCA that maps a high-dimensional vector of statistics from our data to a lower-dimensional vector.

---

9 We thank Raj Chetty and his team for providing us with a mapping between the strings and the SOC codes.
10 For instance, business owners might fill in “self-employed,” which is not a valid SOC code.
11 In Section 3.3, we group individuals into six different categories of employment status based on attachment and type of work.
of moments. Inputs to the PCA are paid- and self-employment income in each sample year and moments of total income averaged across sample years. The specific moments are the mean, the standard deviation, the minimum, and the maximum, with the latter three normalized by dividing by the mean. The number of principal components depends on our choice of the fraction of variance to be explained, which we denote here by $v$. Thus, we have two parameters to choose: the number of neighbors $k$ and the fraction of variance $v$—and we assume they are fixed for all subgroups.

We choose parameters to maximize the predictive accuracy of the $k$-nearest neighbor classifier. To do this, we pull a random sample of subgroups and split them into three subsamples: 70 percent for training, 20 percent for tuning, and 10 percent for validation. For each $(k, v)$ pair, we use the training data to train the classifier and make predictions for the tuning set. We use the validation data to test predictions out of sample. The result of this exercise is $k = 11$ and $v = 75\%$. With these parameters, we apply the classifier to impute skill values for 7 million individuals without usable SOC codes.

2.3.2 Education

The only indicators of education in the IRS microdata are occupation strings with “student” and tuition payment statements (Form 1098-T) filed by eligible educational institutions starting in 1998. To ensure fuller coverage of college attainment, we use a classification algorithm and source data from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) to predict the likelihood of college attainment.

We define individuals as being “college-educated” if they have completed at least an associate’s degree—which would thus include bachelor’s, master’s, professional school, and doctorate degrees. All others are considered “not-college-educated.” For each year $t$, we run the regression

$$\Pr(E_{it} = 1|X_{it}) = \text{CDF}(\beta_t X_{it}),$$

where $E_{it} = 1$ if the individual is college-educated and 0 otherwise for $t$ between 1995 and 2020. The function CDF in (1) is the cumulative distribution function of the standard normal, and variables included in $X_{it}$ are as follows: gender; annual pre-tax wages and salaries; positive business income (equal to 0 if income is negative); negative business income (equal to 0 if income is positive); marital status; number of children (with separate variables for none, one child, and so on, up to nine or more); five-year birth cohort; SOC minor occupation code; and two-digit NAICS industry code. When we used 80 percent of our CPS sample each year to train the classifier and 20 percent to validate the predictions, we were able to correctly predict the education level with 75 to 80 percent

\[12^\text{More precisely, we use an } F_1 \text{ score.}\]

\[13^\text{Some IRS tax filers do not have a valid NAICS code and do not have an SOC minor code. Additional regressions were run using (i) the SOC minor codes with no NAICS; (ii) the SOC major code and NAICS; and (iii) NAICS but no SOC.}\]
accuracy. Coefficients from the CPS-trained classifiers are then used with microdata from the IRS to impute an education indicator for all tax filers in our sample of 25- to 65-year-olds.¹⁴

In Table 1, we report the fraction of individuals categorized as college-educated for the main sample in the first column under “Education and skills.” The imputation results indicate that 53 percent of the main sample is classified as college-educated, which is close to our estimate of the fraction that is cognitively skilled. More individuals are categorized as interpersonally skilled (59 percent) than manually skilled (38 percent).

3 Measuring Returns to Employment

Our main goal is to understand the sources of differences in returns to self- and paid-employment and the implications for theory. To investigate the differences, we start by measuring how income from an activity—either self-employment or paid-employment—varies with age. We first outline the challenges to accurately measure the age profile of income. Then we describe how to use the novel features of our data with an econometric method that is designed to overcome these challenges.

3.1 Some Challenges

A natural starting point for measuring returns to employment is to specify a Mincer-type earnings regression and estimate average income by age after controlling for observables. This regression procedure is widely used and can be implemented with repeated cross sections (see, for example, Hamilton (2000)). Several concerns are associated with this approach. First, differences in average incomes could be driven by selection—we are simply comparing individuals that differ in their latent characteristics. Second, differences in average incomes across age could be driven by the changing composition of the underlying groups.

We develop an econometric approach that addresses these two issues. First, we estimate income by age across activities, allowing for an intercept whose distribution by individual characteristics—whether latent or observed—is essentially unrestricted. Second, we use the long panel aspect of our data to classify individuals based on their attachment to an activity to mitigate problems with composition. We then separately study the income-age profiles of individuals who are attached to their employment status by narrow skills/industry/demographic categories and the income-age profiles of individuals who are less attached and transit into and out of self-employment.

3.2 Econometric Framework

We next describe and motivate the statistical model and estimation procedure that we use to estimate growth in incomes over the life cycle. Our method exploits the presence of multiple

¹⁴All variables in \( X_{it} \) are available in the IRS data, although the IRS occupation field is only available for tax years after 2005 and later (not including 2012) and is a string rather than a SOC code.
cohorts to separately estimate age and time effects for disaggregated subgroups within employment status. For now, we describe the procedure for an arbitrary assignment of individuals to groups and later describe how we construct the groups to minimize selection and composition bias.

We start with some notation. Let $i \in I$ be a set of individuals; $t \in T = \{t_0, t_0 + 1, \ldots, t_0 + T\}$ be a set of calendar dates; $c \in C = \{c_0, c_0 + 1, \ldots, c_0 + C\}$ be a set of birth years (or cohorts); $a \in A = \{a_0, a_0 + 1, \ldots, a_0 + A\}$ be a set of ages; and $g \in G$ be a set of observable time-invariant characteristics (or groups) that partition $I$. Let $y_{i,t}$ be the income of individual $i$ at date $t$. With slight abuse of notation, we use $a(i, t)$ to denote the age of individual $i$ at date $t$, $g(i)$ to denote the group of individual $i$, and $c(i)$ to denote the cohort of individual $i$.

We define two functions $\beta : G \times T \to \mathbb{R}$ and $\gamma : A \times G \times C \to \mathbb{R}$, which capture time, age, and cohort effects. We use the notation $\beta_{g,t}$ and $\gamma_{a,c,g}$ to denote the values of these functions for a particular collection of $\{g, t, a, c\}$, and $\beta_g(i)$ and $\gamma_{a(c(i)), g(i)}$ to be the values associated with an individual-time pair $(i, t)$. Consider the following specification for income:

$$y_{i,t} = \alpha_i + \beta_g(i), t + \sum_{a=a_0}^{a=a(i, t)} \gamma_{a(c(i)), g(i)} + \epsilon_{i,t},$$

(2)

where $\epsilon_{i,t}$ is a disturbance term for individual $i$ at date $t$. The model for income in equation (2) is quite rich. It has three components. First, the parameters $\{\alpha_i\}$ are the unobservable individual-level fixed effects that capture permanent aspects of latent ability, family inputs, and preferences as well as level effects tied to birth cohorts. We impose no restrictions on how these characteristics are distributed in the population or correlated with observable groups. Second, the parameters $\{\beta_{g,t}\}$ are the time effects that vary by calendar time and differ across groups. These parameters capture effects on income such as business cycle fluctuations. Third, the parameters $\{\gamma_{a,c,g}\}$ are the age effects that vary by age, cohort, and group. We are particularly interested in variations across subgroups based on employment status and other characteristics such as skills, industry, and demographics.

It is well-known and easy to see that one cannot separately identify $\beta$ and $\gamma$ from data on income. For instance, for a fixed group $g$, adding a constant to all $\gamma_{a,c,g}$ for which $c + a = t$ is observably indistinguishable from adding the same constant to all $\beta_{g,t}$. To make progress, we impose the following condition.

**Condition 1.** Age effects are the same across cohort bins of size $N_c \geq 2$.

Below, we use the notation $\gamma^0_g$ to indicate the age effect of a group $g$, which is now modified to include a specification of the cohort bin, say, individuals born in the 1950s, 1960s, or 1970s. It is worth pointing out that while we impose the restriction that the age effects for sets of cohorts are the same, we impose no restrictions on how cohorts affect the level of income. The differences
in mean income by cohort are absorbed in the fixed effect for individual \( i \), namely, \( \alpha_i \). Condition 1 allows us to exploit the overlapping structure of our data to separate out age effects from time effects.

Next, we derive the formulas needed to implement the estimation procedure. Let \( \Delta \) be the time difference operator so that \( \Delta x_t = x_t - x_{t-1} \). Apply \( \Delta \) to equation (2) to obtain

\[
\Delta y_{i,t} = \Delta \beta_g(i,t) + \gamma_a(i) + \Delta \epsilon_{i,t}.
\]

We work with differences in levels rather than in logarithms, given that some businesses incur losses and owners’ income \( y_{it} \) can be negative.\(^{15}\) To estimate the age and time effects, we propose the following least squares problem:

\[
\min \{ \Delta \beta_g(t), \gamma_a \} \sum_{g \in G} \sum_{t \in T} \sum_{i \in I} (\Delta y_{i,t} - \Delta \beta_g(i,t) - \gamma_a(i))^2.
\]

By examining the first-order conditions of this minimization problem, we can better understand how the estimator works. Let \( N_{g,t}^a \) be the number of individuals of group \( g \), age \( a \), at calendar date \( t \). Let

\[
\Delta \bar{y}_{g,t} = \frac{\sum_{i \in G} \Delta y_{i,t}}{\sum_{a \in A} N_{g,t}^a}, \quad \Delta \bar{y}_g^a = \frac{\sum_{t \in T} \sum_{i \in G} \Delta y_{a,g}^i}{\sum_{t \in T} N_{g,t}^a}
\]

be the average income growth for group \( g \) between dates \( t-1 \) and \( t \) and the income growth averaged across time for individuals in group \( g \) between ages \( a-1 \) and \( a \), respectively. We can rearrange the optimality conditions to get

\[
\gamma_a^g = \Delta \bar{y}_g^a - \sum_{t \in T} \left( \frac{\sum_{j \in T} \Delta y_{g,t}^j}{\sum_{a \in A} N_{g,t}^a} \right) \left( \Delta \bar{y}_g^a - \sum_{k \in A} \left( \frac{\sum_{t \in T} N_{g,t}^k}{\sum_{a \in A} \sum_{t \in T} N_{g,t}^a} \right) \gamma_k^g \right) \Delta \beta_{g,t}.
\]

Equation (4) expresses \( \{ \gamma_a^g \} \) as linear combinations of two summary statistics of data, \( \{ \Delta \bar{y}_g^a \} \) and \( \{ \Delta \bar{y}_{g,t} \} \) with weights \( \{ N_{g,t}^a \} \). Specifically, the age effects for some age \( a \) are given by the average income growth \( \Delta \bar{y}_g^a \) for that age minus an appropriate weighted average of the time effects \( \{ \Delta \beta_{g,t} \} \). The weights that appear in the adjustment correct for the possibility that the age distribution could be changing over time, which is relevant in our sample period.

To understand the intuition for the adjustment term in (4), consider the case in which the age distribution is constant across time, that is,

\[
\frac{\sum_{a \in A} N_{g,t}^a}{\sum_{a \in A} N_{g,t}^a} = \frac{\overline{N}_g^a}{\sum_{a \in A} \overline{N}_g^a}, \quad \text{(5)}
\]

\(^{15}\)Later we discuss why our econometric procedure avoids issues arising from heteroskedastic errors.
where $N^a_g = \sum_{t \in T} N^a_{g,t}$. With some algebra, we can show that $\bar{\gamma}_g^a = \bar{\Delta} y^a_g - \bar{\Delta} \beta^a_g$, where $\bar{\Delta} \beta_g^a = \sum_{t \in T} \Delta \beta_{g,t} / T$ is the average of time effects for group $g$. It simply says that the estimate of the age effect equals the average income growth for that age minus a simple average of the time effects. However, equation (5) does not hold in typical panel datasets, and therefore the second term on the right-hand side of equation (4) gives the appropriate adjustment.\(^{16}\)

We make two more observations about equation (4). First, the age effect $\bar{\gamma}_g^a$ can be estimated separately for each group $g$. Second, one can show that the rank of the system formed by stacking equation (4) for each age is $A - 1$. Therefore, we need an additional restriction—one for each group—to solve for the age effects $\{\bar{\gamma}_g^a\}$ uniquely. Following Hall (1968) and Deaton (1997), we impose the following condition.

**Condition 2.** The average time effect satisfies

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

for some pre-determined constant $\mu_g$, where $\bar{y}_{g,t_0} = \sum_{i \in I: g(i) = g} y_{i,t_0} / \sum_{a \in A} N^a_{g,t_0}$ is the average income for group $g$ at the beginning of the sample.

Condition 2 allows the estimation to match the cyclical variation in the time effect across groups in a flexible way. This is especially helpful in our sample given the severe economic downturn in 2008–2009. In particular, we do not need to take a stand on the differential effects of aggregate shocks on groups.

### 3.3 Groups

To implement the approach sketched out in the previous section, we need to define groups. A *group* is a Cartesian product of time-invariant characteristics that we call *subgroups*. In our case, there are 46,080 subgroups. In this section, we provide a summary of the subgrouping.

Given our interest in the returns to entrepreneurship, the two most relevant characteristics are: (i) how attached individuals are to market work, whether it is paid- or self-employment, and (ii) how attached a working individual is to self-employment. In Section 2.2, we assigned an employment status to each individual-year observation: “SE” for self-employed, “PE” for paid-employed, and “NE” for non-employed. To address compositional bias, we analyze income profiles by separately studying working individuals who change status and those who do not. We implement that by using the status variable across time to group individuals according to how attached they are to self- or paid-employment. An individual is labeled *attached* if we observe the same employment status for 12 or more tax years with two or fewer changes in employment status during the sample. To be

\(^{16}\)In our sample, we have a balanced panel, and therefore the mean age is necessarily increasing in calendar time as the population is aging.
included in the subsamples of attached self- or paid-employed, we also require that any switching in and out of self- or paid-employment not include intermediate years of non-employment. Those with years of non-employment are categorized separately below.

In Table 1, we report counts and characteristics in the second and third columns for the subsamples of individuals that are attached to paid- and self-employment, respectively. There are 42 million in paid-employment, which is about 52 percent of the individuals. They earn 66 percent of total income, 76 percent of all paid income, and roughly 4 percent of entrepreneurial income. The attached self-employed, numbering 2 million individuals, comprise only 2 percent of the individuals in our main sample but 8 percent of total income and 51 percent of entrepreneurial income.

While small in number, the entrepreneurs in attached self-employment have relatively high incomes when compared to sample totals. Their total combined income from self- plus paid-employment averages $152 thousand—far more than the average attached paid-employed earning $62 thousand. Given our definition of “attached,” it is reassuring to find that most of the total income for self-employed is indeed self-employment income and similarly so for the paid-employed with respect to paid-employment income. Comparing distributions of incomes for these two groups, we find more skewness in self-employment income, as expected. Later, we investigate this further when analyzing the longitudinal data across the life cycle.

Separate results are reported for groups that do more switching in employment status. The almost attached groups have the same employment status for 12 or more tax years but switch more than twice between self- and paid-employment. The shares of the almost attached are much smaller and are analyzed separately as a robustness check, but summary statistics in Table 1 show that they are similar in characteristics to their attached peers. The mostly switchers have 12 or more years in either self- or paid-employment—without an intermediate spell of non-employment—and experience at least 5 or more years in both types of employment. This group is similar in size to the attached self-employed, but those at the top of the distribution earn much less. This group will also be used to gain insight into motivations for entering and exiting self-employment.

The last category is any non-employment that includes individuals that have switched in and out of non-employment from self- or paid-employment at least once or individuals that have five years of non-employment during the sample period. This group is large in counts—roughly 41 percent of the total sample—and as a group accounts for a significant fraction of self-employment income, although they earn on average only $21 thousand in both self- plus paid-employment. If we were to compare the earnings to those of a full-time worker earning the average federal minimum wage for the year, converted to 2012 dollars, we would find that 57 percent have lower total incomes. Even if we condition on those with only one year of non-employment, we find below-average total incomes. These individuals account for 15 percent of the any-non-employed individuals and earn $43 thousand on average—with $35 thousand from paid-employment and $8 thousand from self-
employment. Furthermore, if we condition on the any-non-employed that are more attached to employment, we find that they are few in number and make less on average than those categorized elsewhere. For example, only 28 percent of the any-non-employed do 12 or more years of paid-employment. Only 2.8 percent of the any-non-employed do 12 or more years of self-employment. Those that do self-employment for at least twelve years have on average a total income of $49 thousand in our sample—far less than groups with no intermediate spells of non-employment.

In addition to employment attachment, we use other observables to group individuals. The subgroup College-educated has two values: 1 if the education classifier was above the 0.5 cutoff and 0 if not. The subgroups Cognitive, Interpersonal, and Manual each take on one of two values: 1 if the skill value is above the 0.5 cutoff and 0 if not. Industry is the two-digit NAICS code for the company paying the highest W-2 wages for the paid-employed individual or the company with the highest gross profits owned by a self-employed individual. The industry code takes on 21 possible values (including “missing”). When grouping individuals, we assign them the code observed in most sample years. The subgroup Gender has two values: “M” for male and “F” for female. The subgroup Married has two values: 1 if the individual is married for nine or more years in the sample—not necessarily to the same person—and 0 otherwise. The subgroup Children has two values: 1 if the individual has children and 0 otherwise. The subgroup Cohort has three values: “1950s” if born between 1950 and 1959, “1960s” if born between 1960 and 1969, and “1970s” if born between 1970 and 1975. Since we are working with a balanced panel, we observe a significant overlap of cohorts over time, namely, 26 cohorts (birth years 1950–1975), across 41 ages (25–65) and 16 calendar years (2000–2015).

Summary statistics for education, skills, industries, demographics, and other incomes are included in Table 1 for subgroups with different degrees of employment attachment. Relative to the sample total, all but the group with non-employment spells are more likely to be college-educated and have cognitive and interpersonal skills. There are notable differences in industrial composition across groups. Those with more self-employment attachment are found primarily in construction, professional services, health care, and other services, while the largest sector overall is manufacturing. The demographic data show that 50 percent of the sample is male. Most are married for a majority of years they are in the sample—about 64 percent—and most have children—about 82 percent. The median birth year for our sample is 1963. Across subgroups, we find uniformity in numbers of children and birth year, but some differences in gender shares and marital status.

The two largest groups—attached paid-employed and any-non-employed—have larger shares of women and the mostly unmarried. The any-non-employed were categorized as a separate group.

\[17\]

\[17\] In robustness exercises, we expand \( G \) to include average income deciles in order to check if there are any issues arising from heteroskedastic errors in the baseline regression. We find that our main quantitative results are not sensitive to this change.
in large part because of the many ways they differ from other subpopulations: on average, these individuals earn significantly less, are primarily low-skilled, and have a higher concentration of women than any other category. Not surprisingly, we find average household incomes—whether from spousal wages or asset incomes or transfers such as unemployment insurance—are higher for the any-non-employed than for the employed groups in the sample.\textsuperscript{18}

4 Results

In this section, we report on the estimated entrepreneurial income and growth profiles and then investigate factors affecting occupational choice. Our focus here is primarily on the attached self-employed, but central to our analysis are comparisons to the attached paid-employed with the same demographics, skills, and industries. These comparisons are relevant because the self-employed are rewarded for making firm-specific investments and would thus have potentially different income growth profiles. We also compare growth profiles of the attached self-employed to those for the mostly switching and any-non-employed, two groups that earn a significant fraction of self-employment income in the aggregate. Central to our analysis of occupational choice are comparisons of past wage, asset, and other household incomes for those that enter self-employment and those that do not. We investigate these choices for the full sample and for the youngest cohorts that experiment most. Finally, we relate our findings to the existing empirical literature and emphasize results most relevant for distinguishing between theories of occupational choice.

4.1 Entrepreneurial Incomes

We start with the main findings of the least squares estimation of equation (3). Estimates of the time effects provide a summary of cyclical growth patterns, including impacts during the 2008–2009 downturn. Estimates of the age effects provide a summary of life-cycle growth patterns for the self-employed, which can be compared to more familiar patterns of the paid-employed. Netting out time and age effects, we then analyze the residual growth, which contributes most to the variability of incomes and is thus relevant to understanding the nature of risk-taking in entrepreneurship.

4.1.1 Cyclical Growth

In Figure 1, we plot the time effects relative to average income for individuals that are attached, to either self- or paid-employment, that is, a weighted sum of $\Delta \beta_{g,t}$ divided by $\bar{y}_{g,t}$ for each subgroup $g \in \mathcal{G}$ where $\mathcal{G}$ is either the attached self- or paid-employed group and weights are constructed

\textsuperscript{18}Asset income is the sum of the following categories reported on Form 1040: taxable interest, tax-exempt interest on municipal bonds, dividends, Schedule D profit or loss, and Schedule E profit or loss (other than distributions from S corporations and partnerships).
with subgroup counts. The values are reported in percentages and displayed for tax years 2001 to 2015. As expected, a decline in growth occurs during the Great Recession, with paid-employed growth rates falling to a low of −1 percent and self-employed growth rates falling to a low of −16 percent. As we see from the figure, the flexibility of the econometric specification allows for differences in timing and magnitudes. Declines start earlier for the self-employed and are much larger in the midst of the downturn than those for the paid-employed. Interestingly, both groups see improvements by 2010, continuing on until 2012.

4.1.2 Life-cycle Growth

In Figure 2, we display the integrated incomes and associated growth profiles for the same sample of attached self- and paid-employed used in Figure 1. Panel A shows the weighted average integrated incomes. For each age \(a \geq 25\), we compute average income for 25-year-olds in group \(g\) plus \(\sum_{j=26}^{a}(\gamma_j^{g} + \Delta\beta^{g})\) and denote this sum for group \(g\) at age \(a\) by \(Y^{g}(a)\). In Figure 2, we plot the averages using sample counts for weights, that is, \(\sum_{g}N^{g}Y^{g}(a)/\sum_{g}N^{g}\) for both groups. In the figures, dots are the point estimates and bold lines are the third-order polynomial fits. Panel B shows the weighted average growth by age, that is, \(\{\gamma^{g} + \Delta\beta^{g}\}\) where weights are again based on group counts. We also report the integrated incomes at age 25 and 55 from Panel A, as we do in growth profile figures for subpopulations shown later. As the figure shows, the growth profiles differ substantially across the groups. The self-employed incomes are significantly higher by age 55 relative to their paid-employed counterparts. Another striking difference is the life-cycle growth patterns shown in Panel B. Growth rates in paid-employment decline across the life cycle, whereas growth rates in self-employment do not. The attached self-employed have persistently high average growth in incomes—in the range of $7 thousand to $9 thousand annually—for ages between 25 and 40, whereas the paid have less than $4 thousand in all years. Even after age 40, the self-employed average growth rate remains significantly higher than that for paid-employees and, at age 55, the self-employed have an average income of $210 thousand—more than twice that of the paid-employed group and too large to be attributable to differences in taxes or benefits.

Growth Profiles for Attached Subgroups. Next, we repeat the exercise for subgroups of the attached self- and paid-employed, specifically by gender, marital status, education, skills, and industry. In Figure 3, we report growth profiles for men in Panel A and for women in Panel B. Recall that men account for most of the attached self-employment sample, roughly 82 percent, but only 53 percent of the attached paid-employment sample. Thus, it should not be a surprise that the self-employed growth pattern for men is nearly the same in magnitude and shape as the full sample. Perhaps more surprising is the growth pattern for self-employed women, which is also higher than

\[19^{19}\]In all results reported, we exclude the top and bottom 0.01 percent outliers. Our main aggregated results are not affected if we include them.
for their paid-employed counterparts. For the paid-employed, neither the men nor women show any increase in growth at age 25. When integrated, the income profiles reveal large level differences between paid- and self-employment for both men and women by age 55. However, we find a gender gap when we compare peak incomes, regardless of employment status. Average income for age 55 self-employed women is $137 thousand—or 61 percent of that for men. Average income for age 55 paid-employed women is $72 thousand—or 71 percent of that for men.

Another common attribute for the attached self-employed is being married in most years of the sample. Figure 4 reports growth profiles of the attached sample by marital status and incomes at age 25 and 55. Here again, we find similar results when comparing the mostly married to the full sample since they account for roughly 79 percent of the attached self-employed population and 68 percent of the attached paid-employed. The mostly unmarried self- and paid-employed have similar qualitative patterns in growth, but the incomes of the self-employed are $58 thousand higher than those of the paid-employed by age 55.

Education is another important characteristic when considering returns to employment. After applying our classifier to categorize individuals as “likely” or “not likely” to have attained a college degree, we recompute slopes and construct average growth profiles. The results shown in Figure 5 reveal large differences in outcomes for those classified as college-educated when comparing across employment status and, in the case of the self-employed, when comparing across education levels. By age 55, we predict that the average income for college-educated self-employed individuals is $321 thousand and only $111 thousand for their paid-employed peers. The difference is even larger if we compare these self-employed individuals with those classified as not college-educated. Average income at age 55 for the latter group is equal to $42 thousand. Interestingly, as the figure shows, there is hardly any difference in outcomes if we compare self- to paid-employed if they are classified as not college-educated.

With the O*NET data and our occupation strings, we can study returns to self- or paid-employment for people with different skill sets. In Figure 6, we show growth profiles by cognitive skill. A comparison between this figure and that for education (Figure 5) reveals some difference in the categorizations. For example, differences for those classified as college-educated and not-college-educated are much starker than differences between cognitively skilled and not cognitively skilled—regardless of occupation choice. In the case of the self-employed, those classified as cognitively skilled have an average income at age 55 of $222 thousand, while the not cognitively skilled have an average income of $192 thousand.

In Figure 7, we show growth profiles by interpersonal skill. While not as stark as our comparisons by education, there are still large differences between the interpersonally skilled self- and paid-employed and between the skilled and unskilled regardless of employment status. For example, as with the college versus non-college comparisons, we again find a much more pronounced
hump-shaped growth profile for the interpersonally skilled self-employed when compared to the interpersonally unskilled self-employed. We again find that the non-interpersonally skilled have much lower growth profiles and average incomes by age 55 when compared to their skilled peers, although not as low in magnitude as for the non-college-educated shown in Panel B of Figure 5. In Figure 8, we show growth profiles by manual skill. We find the patterns for individuals working in occupations with manual skills are similar to those described for the non-college-educated.

Another relevant cut of the data is by industry since the self-employed tend to be clustered in particular occupations and sectors. In Figure 9, we plot results for individuals with attached employment status that are in professional services (NAICS 54) and health care (NAICS 62). Together these sectors account for 27 percent of the attached self-employed population. Here, we see hump-shaped growth profiles for entrepreneurs and large differences between the income growth rates of the self- and paid-employed. By age 55, the attached self-employed average incomes in professional services and health care are $304 thousand and $283 thousand, respectively, whereas the average across all industries is $210 thousand.

The results shown thus far are for relatively broad categories and include people with a wide range of characteristics. Because the data have thousands of possible cuts—even if we condition on being in the attached groups—plotting all of them is not possible. However, we are able to highlight the most important groups by ranking them according to their importance in generating differences in the average income growth rate for the attached self-employed and the attached paid-employed. As shown in Figure 2, there is a sizable gap in income growth between ages 30 and 39—with an average at roughly $5 thousand in 2012 dollars. In Table 2, we summarize the groups that make up at least 50 percent of this difference. The first column summarizes the cumulative share. Reading across the row, we report distinguishing characteristics of the groups. We do not list the characteristics that they all share. It turns out that all top contributors are male, mostly married, college-educated, interpersonally skilled, and have kids. In terms of the distinguishing characteristics, we find that the group contributing the most to the growth differential works in the health care sector (NAICS 62), is born in the 1960s, is cognitively skilled, but is not manually skilled. This group of primarily medical doctors contributes 10 percent to the growth gap, with the self-employed among them experiencing average annual growth in income around $17 thousand during their 30s—roughly 1.6 times the growth of their paid-employed colleagues. As Table 2 makes clear, a small number of sectors matter for our results: health care, professional services, construction, finance, and retail trade.

**Growth Profiles for Non-Attached Groups.** Thus far, we have compared the attached self-employed to their paid-employed peers. Since a significant fraction of self-employed income is earned by our mostly switchers and any-non-employed subgroups, we also include a comparison with these
less-attached peers. In Figure 10, we compare the growth profiles for the latter groups to that of the attached self-employed. Panel A shows the growth profile comparison between the attached self-employed and the mostly switching. This figure highlights the delay in growth for those more attached to self-employment. Also interesting is the fact that the growth of the switchers is higher at age 25 but is declining in most years over the life cycle, reminiscent of the paid-employed growth patterns shown earlier.

Panel B shows the growth profile comparison between the attached self-employed and the any-non-employed.20 Individuals categorized as any-non-employed have little self-employment income on average, but when aggregated, the sum is about 20 percent of all self-employment income. However, the growth profile for the any-non-employed has the same pattern as that of the paid-employed—just lower at all ages. This finding may not be too surprising given these individuals are less educated, lower-skilled, and likely to be second earners. But more surprising is the fact that many of these individuals are out of employment only one year and still have very low incomes and growth rates.

4.1.3 Residual Growth

We turn next to an investigation of the variability and persistence of income growth, which is relevant for discussions of entrepreneurial risk-taking and, more generally, for the larger debate on earnings inequality. We show that critical to this investigation is the estimated residual growth—that is, \( \Delta \epsilon_{i,t} \), found after netting out the time and age effects from the total income changes—as this term accounts for most of the variation in overall growth. After reporting on key statistics of changes in overall and residual incomes, we work through a simple analytical example to demonstrate how these moments of the data can be used to estimate the potential gains of insuring income risk.

We start by computing year-to-year transitions in income levels, \( y_{it} \), as a point of reference for both the self- and paid-employed. These matrices are reported in Table 3, with the results for the self-employed in Panel A and those for the paid-employed in Panel B. Each element of the matrix is the share of individuals that start the year in a particular income bin (listed in the first column) and transit to one of the income bins listed at the top of the matrix in the next year. Along the bottom of each matrix are the distributions of individuals in each bin. We have purposely chosen the positive income bins to have the same log-spacing between bins with 1 percent of the self-attached sample in the top bin. Earners in this bin have total incomes—from self- and paid-employment—above $1.6 million.21 We also include a separate bin for negative incomes, which is

---

20 When computing averages for the any-non-employed, we only include observations with non-zero incomes in both the previous and current year.

21 We also computed income transitions for the attached self-employed using deciles. DeBacker, Panousi, and Ramnath (2022) report transition matrices with business income deciles, but given differences in sampling choices,
possible for both the self- and paid-employed, given we are reporting their total incomes that could include self-employment losses. Comparing rows across the two transitions, we find more dispersion for the self-employed in middle incomes when compared to the paid-employed, but less so at the top and bottom.

To provide a sense of the potentially risky nature of entrepreneurship across the life cycle, we compute percentiles of growth rates by age, that is, $\frac{\Delta y_{i,a}/|y_{i,a-1}|}{1}$ between ages $a - 1$ and $a$ for primary working ages 26 to 55 (and thus avoid later years when individuals are likely to work less or retire). Because incomes are both positive and negative, we divided the income changes by the absolute value of income in age $a - 1$. The percentiles of income growth are plotted in Figure 11, with results for the attached self-employed in Panel A and results for the attached paid-employed in Panel B. For both groups, the income changes are most dispersed at younger ages. As expected, the self-employed incomes show more dispersion in growth rates at all ages. However, the 90–10 variation is relatively constant across middle ages for both groups, suggesting that the volatility in self-employed incomes is not rising over the life cycle relative to the volatility in paid-employed incomes even though differences in average incomes are rising over the life cycle. This feature is relevant for theories that would rely on risk compensation as an explanation for the differences in mean growth rates between paid- and self-employed individuals.

The data underlying Figure 11 can also be used to compute the variability of income changes within and across subgroups of our sample. Knowing this allows us to determine whether individuals in different subgroups—for example, those in different industries—face significant differences in income risk. Knowing this allows us to determine if there are significant differences in income risk faced by individuals in different subgroups, say for example, by those in different industries. The growth in observed incomes by age may be computed in two ways: we can pool the populations of attached self- and paid-employed and plot the dispersion, as in Figure 11, or we can compute the statistics for subgroups of these populations and construct weighted averages using population counts for weights. Although not shown, we find that the dispersion in income changes is nearly the same regardless of whether we pool the individuals into attached self- or paid-employed categories or instead compute a weighted average of the underlying subgroups. For both methods of aggregating, we find results similar to what is shown in Figure 11. This means that the within-group variation of income changes accounts for nearly all of the variation.

With time and age effects netted off, our estimation procedure yields residual growth for all individuals $i$ in tax year $t$, namely, $\{\Delta \epsilon_{i,t}\}$, which is of independent interest when modeling shocks to income. As we show later, two features of these data are particularly useful: the dispersion

---

their results are not directly comparable to ours. These authors only drop filers who never report business income outside of $[-5000, 5000]$ in 2012 dollars. Forty percent of the observations in their sample are included in bins below $5,000$—many more than we find for our attached self-employed sample. Furthermore, the lower cutoff for the top decile in their sample starts at $82,000$, whereas our cutoff when reporting deciles is $333$ thousand.

21
and autocorrelation. Figure 12 plots the percentiles of residual growth over the life cycle after first normalizing the age \( a - 1 \) to age \( a \) change by the absolute value of lagged income \( |y_{i,a-1}| \). As shown, the results in Figure 12 reinforce the earlier evidence in Figure 11 and again show that the volatility in income changes is decreasing with age for both the self- and paid-employed. In Table 4, we compute the transition matrices for these data. As before, we report the share of individuals that start in a particular bin listed in the first column and transit to one of the income change bins listed at the top of the matrix in the next year. Distributions of these income changes are shown at the bottom of each matrix. These matrices can be used to infer how autocorrelated the income changes are and, in turn, how persistent or temporary they are. As we see from the results for the attached self-employed, the probabilities are high in the upper right and lower left of the matrix, indicative of a negatively autocorrelated process. The more negative the autocorrelation, the more temporary the income change. We see the same pattern for the paid-employed income changes reported in lower panel, which suggests that the main difference for the two groups is in the dispersion, not the persistence, of income changes.

The moments of the data discussed thus far can serve as useful inputs to welfare calculations of the gains to fully insuring against idiosyncratic risk. To demonstrate how this is done, we work through a concrete example, making specific assumptions about the processes governing income growth—or, in our case, residual income growth—and the risk preferences of our individuals in order to derive analytical relations between data moments and welfare. Suppose that (i) growth rates in income are well summarized as the sum of a nonstationary random walk process \( r_t \) and a stationary autoregressive process \( z_t \) and (ii) preferences are of the Epstein and Zin (1989) class. To make analytical progress, we need a few more assumptions: the autoregressive process is not serially correlated, the shock processes are Gaussian, the intertemporal elasticity of substitution parameter in preferences is equal to 1, and consumption moves one-for-one with permanent shocks to income and not at all to transitory shocks, as dictated by the permanent income hypothesis. Then, with simple algebra, it is straightforward to show that the fraction of wealth \( \lambda \) an individual would forgo to fully insure their risky income is given by

\[
\lambda = -\frac{1}{2} \alpha \beta \sigma_r^2,
\]

where \( \sigma_r^2 \) is the variance of the Gaussian shocks of the random walk process and \( \alpha \) and \( \beta \) are parameter inputs to the utility function \( V \) over paths of consumption \( \{c_j\} \) proposed by Epstein and Zin (1989), namely,

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

22 With the aid of computer simulations, it is straightforward to relax these assumptions and consider more general specifications.
Equation (7) is derived by taking the limit as $\rho$ approaches 0, consistent with an intertemporal elasticity of substitution equal to 1.

To make progress quantitatively, we can use estimates of 90–10 differences and autocorrelations for income changes to infer the variance $\sigma^2_r$ in (7). With these moments and the assumptions given above, we can derive an analytical relation between the IRS statistics and predictions for variances of the permanent and temporary shocks underlying the income changes. Let $Q$ be the 90–10 difference in the income changes, and let $A$ be the autocorrelation. Then, we can show that

$$Q = 2.56\sqrt{\sigma^2_r + 2\sigma^2_z}$$

$$A = \frac{\sigma^2_z}{\sigma^2_r + 2\sigma^2_z},$$

where the 2.56 is equal to the 90–10 difference for a standard normal distribution and $\sigma^2_r + 2\sigma^2_z$ is the total variance of the income change if the temporary shock $z_t$ is an independent and identically distributed random variable.\(^{23}\) Note that the variance of the temporary shock is multiplied by two since we are analyzing the change in income. From equations (9)-(10), we derive an estimate of the variance for this example:

$$\sigma^2_r = \left(\frac{Q}{2.56}\right)^2 (1 + 2A).$$

In the case of the residual growth process, estimates for the 90–10 differences can be read right off Figure 12. If we average over the life cycle, we find $Q$ around 1.61. Using Table 4 for the self-employed, we find $A$ to be roughly $-0.16$. Using the formula above, we then have an estimate for the permanent shock standard deviation $\sigma_r$ of 0.52 and an estimate of the temporary shock standard deviation $\sigma_z$ of 0.25. Suppose we set the discount factor $\beta$ in the Epstein-Zin preferences equal to 0.96 and the risk aversion parameter $\alpha$ equal to $-1$. In this case, we would predict that the fraction of wealth ($\lambda$) an individual would forgo to fully insure their risky income is 12.8 percent. If we repeat the exercise for the paid-employed and assume the same preference parameters, we find that $\lambda$ is around 1.4 percent. In this case, differences in $\lambda$ are driven primarily by differences in dispersion (which has a ratio around 3) because the autocorrelations are not that different between the two groups ($-0.16$ versus $-0.17$).\(^{24}\)

While the exercise of translating the data moments into measures such as $\lambda$ provides economically interpretable summaries of the risk entrepreneurs face, it is important to keep in mind that these calculations assume a one-for-one pass-through of income shocks to consumption, which allows analytical tractability but is too extreme. We know from Table 1 that 79 percent of these

\(^{23}\)We derive these results along with equation (7) in a separate appendix.

\(^{24}\)Abowd and Card (1989) report an average autocorrelation of $-0.29$ for changes in logged earnings of paid-employed male household heads in the Panel Survey of Income Dynamics (excluding the Survey of Economic Opportunity subsample), with the range of estimates for 1969–1979 given by $[-0.54, -0.10]$. We find an average autocorrelation of $\Delta \log y_{it}$ for the paid-employed equal to $-0.25$, regardless of whether we restrict attention to males or married males or include all paid-employed.
individuals are married most years of our sample. As we report at the end of Table 1, the average wage of their spouses is $26 thousand before accounting for any employee benefits. We also know that the average household asset income for this group—including payments of dividends, interest, and capital gains—is $58 thousand, implying sufficient saving stocks in downturns.

More generally, if the total adjusted gross income (AGI) is a better proxy of what household members have available for consumption, then a more quantitatively relevant pass-through rate is the change in AGI following a change in self-employment income. To compute this, we estimate equation (2) with $y_{it}$ set equal to AGI and construct new residual growth rates. We then regress these rates on those computed above (with $y_{it}$ equal to the income of attached self-employed individuals). When we do this, we find a coefficient from the regression equal to 0.16—well below 1 assumed in the calculations above. In other words, the analytical calculations above, while economically interpretable, should be viewed as an upper bound on the welfare gains for fully insuring consumption, as individuals in our sample already partially do through other means.

4.2 Entrepreneurial Choice

We turn next to analyzing individuals who switch across employment status. Understanding switching behavior is key for theories of occupational choice. In this section, we measure the extent of switching and analyze differences between those who switch and those who do not. We find that entry rates into self-employment are relatively flat across the life cycle and across time. Exit rates out of self-employment decline with age but vary little over time, even during the 2008–2009 recessionary period. Importantly for theory, we find that the switching behavior reveals positive selection on past incomes, negative selection on asset and spousal wage income, and roles for both non-pecuniary and pecuniary motives driving entry in and exit out of self-employment.

We focus on switching rates between employment activities, namely, paid-, self-, and non-employment. For a group of individuals, a switching rate from activity $A$ to $B$ is defined as the fraction of individuals whose status was $A$ at age $a$ (or date $t$) and $B$ at age $a + 1$ (or date $t + 1$). The entry rate into activity $A$ is the fraction of individuals who transit from not-$A$ at age $a$ (or date $t$) to $A$ at age $a + 1$ (or date $t + 1$), and the exit rate is defined analogously.

In Panel A of Figure 13, we plot the entry rates into self-employment from either paid- or non-employment or both by age. The figure shows that the overall entry rate is in the range of 1 percent to 2 percent and is modestly increasing in age. Although not shown here, there is also a distinct gender gap in the overall entry rate, with women’s rates significantly lower. As is clear from the figure, most of the rise is due to entry due to non-employment. If we condition on gender,

---

25 Before running the regression, we drop the top and bottom 0.01 percent outliers for both residual growth rates.
26 Entrepreneurs also have outside opportunities that provide additional insurance, such as switching to paid-employment, declaring bankruptcy, or listing their firms. See, for example, Manso (2016) and Catherine (2022).
we find that most entry from non-employment is actually men and not simply a return of women to the ranks of the employed after having children.

Exit rates out of self-employment are shown in Panel B. The overall rate is high and strongly declining, starting around 38 percent and dropping to about 17 percent by the end. If we compute these statistics for women and men, we again see a gender gap: relative to men, women have exit rates that are roughly 6 percent higher at all ages. The declining hazard rate—whether for men or women—is suggestive that experimentation and learning about the potential gains to entrepreneurship occur early in careers. Most of those switching at early ages go into paid-employment. Not surprisingly, by the end of the life cycle, more switch to non-employment because of early retirements.

In Figure 14, we plot the entry and exit rates by tax year. In this case, we purge an age effect that arises from the aging population over our sample period.\textsuperscript{27} We find that, between 2001 and 2015, the entry and exit rates are remarkably flat with no clear time trend. The lack of cyclical variation around 2008–2009 suggests that self-employment is not used by many as a hedge against unemployment risk (see, for example, Alba-Ramirez (1994), Evans and Leighton (1989), Rissman (2003), and Rissman (2007)).

To better understand the motives and impediments to switching, we compare past incomes for individuals who switch as some age to that of comparable individuals who switch as some later age.\textsuperscript{28} Our first exercise compares the past wage income for one-time switchers into self-employment to the past wage income of those who switch later but share the birth year, gender, industry, marital status, and lagged employment status, whether it be paid- or non-employment.\textsuperscript{29} We use three years of data when computing past incomes and when comparing lagged employment status. More specifically, let $x_{i,t}$ be the variable used to predict the switch—say, past wage income in this case—for individual $i$ at time $t$, and let $x_{m(i),t}$ be the same variable for all matched peers $m(i)$ of individual $i$. Then we compute the difference $\Delta_{it}$ in the averages of variable $x$ before the switch as

$$\Delta_{it} = \frac{1}{3} \left[ \sum_{j=1}^{3} x_{i,t-j} - \frac{1}{N_{m(i)}} \sum_{m(i)} \sum_{j=1}^{3} x_{m(i),t-j} \right].$$

In Panel A of Figure 15, we plot the interquartiles of this difference in past income by age of switch. A positive value indicates switchers have higher past income than future switchers. We see that early switchers have similar past incomes to non-switchers, and over time the gap becomes larger.

\textsuperscript{27}Specifically, after constructing switching rates for each date $t$, we subtract a weighted difference of the age-$a$ switching rates, with weights equal to the age-$a$ share of the population in $t$ less the age-$a$ share of the population in year 2001.

\textsuperscript{28}We also compare results for switchers to those of comparable non-switchers and find quantitatively similar results for all panels of Figure 15.

\textsuperscript{29}In these comparisons, we use yearly indicators of married or not married rather than the time-invariant notion of “mostly” married.
and more favorable for the switchers. By age 55, the median difference is roughly $2 thousand. These findings hold up even if we focus exclusively on those in paid-employment prior to the switch. From this exercise, we conclude that most switchers are positively selected on past productivity.

Next, we compare asset incomes of switchers and future switchers. Since labor and asset incomes are generally correlated, we isolate the role of assets for entrepreneurial choice by comparing switchers to later-switching peers who not only share birth year, gender, industry, marital status, and lagged employment status, but also the percentile of past wage income. In Panel B of Figure 15, we plot the distribution over age of past asset income of the switchers less an average of past asset income of the future switchers they are paired with. Call this difference the excess asset income. For most switchers, we find the differences to be negative and small. As in the case of the past wage income comparisons, the differences in incomes are small for younger ages and grow larger (in absolute value) for older ages. By age 55, the median difference in asset incomes is −$2 thousand, with the current switchers earning less than the future switchers. From this exercise, we conclude that most switchers are negatively selected on liquidity.

Panels C and D of Figure 15 repeat the exercise but in these panels, we use past spousal wages and adjusted gross incomes instead of asset incomes. Like asset incomes, other incomes earned by the household can be used for financing business startups or for smoothing consumption in the early years of operations. In the case of spousal wages, we find the differences to be small or negative for most of the distribution. Not surprisingly, in the case of adjusted gross income—which is a much broader measure of income—we find larger differences across the distribution when comparing switchers and future switchers. However, the differences here are roughly split between positive and negative and, thus, there is no definitive conclusion about selection.

In order to investigate pecuniary versus non-pecuniary motives for switching, we compute the change in income following a switch in employment status from self- to paid-employment or vice versa. Declines in income would be consistent with a potential role for non-pecuniary motives. To test this, we compare the average incomes of individuals that switch their employment status—averaging over the incomes in the three years subsequent to the switch—and compare this income to the three-year averaged income of a control group that did not switch. As above, we assume the control group has the same birth year, gender, industry, marital status, and three-year lagged employment status as the switcher but remains in the same employment status for at least three more years. We also condition on the skill set—which is a good proxy for occupation—and the fact that individuals in the control group do switch their status at some point outside of the six-year window. In Panel A of Figure 16, we plot the income differential following a switch—along with the interquartile ranges—weighted by subgroup counts, for ages 28 to 60. We find the median income change post-switch is negative but small. Overall, we find a large fraction of the population

\[30\text{ For these results, we exclude Schedule D capital gains from asset income.}\]
of switchers earning more post-switch and a large fraction earning less than non-switching peers, suggesting that both non-pecuniary and pecuniary motives are driving entrepreneurship.

Results for those switching from self- to paid-employment are shown in Panel B of Figure 16. Here again, we find that the results are split: there are almost as many with higher post-switch income as there are with lower. The exceptions are the youngest entrepreneurs that see more pay gains than losses when switching to paid-employment. We turn to examining this group in more depth next.

4.3 Young Entrepreneurs

Our results thus far suggest that entrepreneurs under the age of 40 have higher exit rates when compared to older peers and, at least for those that stay, have steeper growth profiles and more volatile incomes. This subgroup of our sample is particularly interesting because young business owners are less likely than older peers to have previous self-employment experiences or opportunities to accumulate much in the way of assets before starting a business. With this in mind, we revisit some themes already discussed above but with a narrower focus on the youngest cohorts born between 1970 and 1975.

To investigate the high exit rates of young entrepreneurs, we ask if there are differences between young individuals who experiment with self-employment while young and continue on in business with those that experiment but then exit. We track individuals in the youngest cohorts with at least five years of self-employment experience prior to age 35. In Figure 17, we report the growth profiles for those that continue in self-employment after age 35 to those that switch to paid-employment. The figure reveals a familiar pattern: the growth profile of those continuing in self-employment is higher and more hump-shaped than the profile of those who switch into paid-employment. Those that stay have profiles more similar to attached self-employed, and those that exit have profiles more similar to attached paid-employed. One explanation is that the switchers were never committed to the entrepreneurial path in the first place and did not make the necessary firm-specific investments. Another explanation is that the switchers learned early that they have low entrepreneurial skill and exited. Later, we test these hypotheses.

In terms of volatility, a potential issue for young entrepreneurs—especially those starting new businesses—is financing start-up costs and early firm-specific investments. Here, we characterize the path of initial losses at both the business level and individual level for a group of business founders that are in one of the 1970–1975 cohorts. We define founders to be individuals filing a Schedule K-1 as an owner in the first year that an S corporation or partnership starts. We further restrict attention to businesses that have at least eight years of consecutive tax filings with business receipts or deductions. For this sample, we find that the business net income is negative for 45 percent in the first year of business, 35 percent in the second, and 32 percent in the third. Flipping
this around, we ask when these businesses have a first positive net income. In this case, we find 53 percent of businesses in the first year, 19 percent in the second, and 8 percent by the third.

Interestingly, if we ask these questions of the new owners in terms of their own self- plus paid-employment income, we find a relatively low pass-through of losses. In the first three years of business, we find that the owner’s total income (from self- plus paid-employment) is negative for 10 percent of owners in the first year, 9 percent in the second, and 8 percent in the third. If we ask when these owners have their first positive total income, we find 90 percent in the first year, 5 percent in the second, and 2 percent in the third. The higher share of owners with positive income at an earlier stage is further evidence that self-employed individuals have other means of smoothing personal consumption expenditures over time—say, because they have multiple businesses or because they supplement the early year business incomes with income from paid-employment.

4.4 Comparison to existing literature

To better motivate the theory that we develop next, we first relate our empirical findings to those in the existing literature. A large literature uses survey data for the United States to investigate entrepreneurial income profiles and occupational choice. Prominent examples are Lazear and Moore (1984) with the Current Population Survey, Evans and Leighton (1989) with National Longitudinal Survey of Youth (NLSY), Hamilton (2000) with Survey of Income and Program Participation, Hurst and Lusardi (2004) with the Panel Study of Income Dynamics, and Moskowitz and Vissing-Jorgensen (2002) with the Survey of Consumer Finances.\(^{31}\) This literature has been extremely influential in promulgating our understanding of entrepreneurship and motivating theories that can be used for policy analysis. In this section, we relate our findings to these studies—delineating points of both agreement and disagreement.

Since survey data have issues related to top-coding and small samples, most research on entrepreneurship has focused on the median incomes of the self-employed. To relate our findings to those based on surveys, we start by comparing simple cross-sectional moments—medians and means—for self-employment income and paid-employment income based on data from the CPS and IRS. The IRS sample used for this comparison is different from that summarized in Table 1 to ensure consistency with the CPS. More specifically, we use data for all available cohorts and only two criteria for the self-employment assignment: the absolute value of income must exceed $5,000 (in 2012 U.S. dollars) and must be greater than the income from paid-employment.\(^{32}\) In Panel A of Figure 18, we plot median self-employment incomes by age for the IRS and CPS.\(^{33}\) While the IRS profile is lower and steeper at early ages than that based on CPS data, the series show median

---

\(^{31}\)For a comprehensive set of references, see Parker (2018).

\(^{32}\)As before, we exclude the top and bottom 0.01 percent outliers. Including these individuals adds more noise to the series of cross-sectional means but barely changes the polynomial fit.

\(^{33}\)Quantitative results shown in Figure 18 are robust to using the main sample of analysis underlying Table 1.
incomes that are close between the ages 45 and 60. If we compare either series to the paid-employed counterparts in Panel B of Figure 18—where we find almost no difference between the CPS and IRS data before age 55—we recover a familiar result: median self-employment income is below median paid-employment income. These results are consistent with an abundance of survey evidence that finds a self-employment “discount.” This finding has solidified the view that self-employed individuals must be earning large non-pecuniary benefits from being their own boss and having flexible jobs (see, for example, Hurst and Pugsley (2011) and Catherine (2022)). Similar conclusions are drawn by Moskowitz and Vissing-Jorgensen (2002), who emphasize low returns relative to the risk in self-employment.

In the lower panels of Figure 18, we plot cross-sectional means for each age. Panel C has results for self-employment incomes reported in the CPS and the full IRS sample, and Panel D has analogous results for paid-employment incomes. Two comparisons should be made here. First, we need to compare survey results against the IRS population. In the case of self-employment income, the differences are large: the IRS average income is close to $42 thousand higher than the CPS average at the peak. In the case of paid-employment, the differences are small. The second comparison is across employment status. Using the IRS data, we would conclude that the self-employed earn significantly more than the paid-employed. Using the CPS data, we would conclude that the paid-employed—in particular, those in prime working ages—earn only modestly more than the self-employed.

The CPS-IRS comparisons across the means and medians suggest that the discrepancies are driven by the properties of the right tail. To further investigate this, we compare the 75th, 90th, and 95th percentiles and shares of incomes above those percentiles in comparable CPS-IRS groups. For self-employed individuals, the differences are stark. The right tails in the CPS are significantly thinner than their IRS counterparts.

The fact that the self-employment income distribution is right-skewed means that the typical dollar in self-employment does not come from the typical self-employed individual. To further

---

34 Although Levine and Rubinstein (2017) claim that median incomes are higher for the incorporated self-employed, their estimate of the difference is only $5,000 above the paid-employed for the NLSY. This finding is consistent with Hamilton (2000), who documents smaller differences between the incomes of the self- and paid-employed at higher quantiles of the distribution where the incorporated owners would naturally be.

35 Hall and Woodward (2010) analyze data for the universe of U.S.-based high-tech start-ups and find expected returns for the owners are higher than from salaried work, but also conclude that the returns are roughly equal after taking into account the risk owners face.

36 This difference does not correct potential underreporting in the tax data. For instance, for a matched CPS-IRS sample of self-employed individuals, Imboden, Voorheis, and Weber (2022) find that CPS respondents report 51 percent less income to the IRS.

37 Bollinger et al. (2019) compare average CPS and SSA W-2 earnings for 440 thousand individuals that appear in both samples. Over the period 2005–2010, they find the difference in average income is $813 (reported in 2010 dollars).

38 The econometric approach laid out in Section 3.2 exploits the long panel that administrative data provides to separately analyze groups that differ with respect to their attachment to self-employment. This is the main advantage over working with simple averages.
explore this, we introduce a new group based on ranking individuals by their average income, once we have conditioned on their NAICS code, cohort, and gender. To ensure that we compare slopes of income profiles by employment status for individuals with similar average incomes, we deliberately ignore their employment status (paid- versus self-employment) before assigning them a rank. After ranking them, we bin individuals into five quintiles. Table 5 shows the shares of income after individuals have been ranked for total, paid-, and self-employed income. In the case of self-employed income, we see that 89 percent of the self-employed income comes from individuals in the top 25 percentiles, and a majority of the latter are those we classified as attached self-employed.

In Figure 19, we plot the growth profiles for those in the top 25 ranks in the case of the attached self- and paid-employed (Panel A) and those in the bottom 25 ranks (Panel B). The figures show that for the top 25 ranks the differences in growth rate by age are starkly different across self- and paid-employed, whereas for the bottom 25 ranks they are not. From this, our takeaway is that most self-employment income is characterized by the patterns we highlighted in Section 4.1, that is, with steeper, more persistent income growth for the self-employed as compared to the paid-employed. In other words, we find patterns for the top ranks that are quite different from those emphasized by the current literature. For the bottom ranks, on the other hand, we see that the self- and paid-employed patterns are similar, as we saw earlier in subgroups such as the non-college-educated. However, these groups have only a small share of either paid- or self-employment income.

Next, we compare our findings on entrepreneurial entry and exit. As far as switching rates by age are concerned, our estimates are in line with those from surveys (see, for example, Evans and Leighton (1989) and Fairlie (2005)). Our findings that entry and exit rates do not show a trend or fluctuate much around the 2008–2009 recession might seem contradictory to the findings from U.S. Census data, such as Decker et al. (2014), which shows a decline in the start-up rate of around 25 percent at the start of the Great Recession. However, their findings largely reflect differences in samples. Studies that find declining entry rates use measures such as the fraction of new firms in the Longitudinal Business Dynamics (LBD) data, whereas we focus on individuals who enter self-employment. To reconcile the differences, we analyze a smaller sample of self-employed that is more aligned with the LBD, namely, those with employees. This group includes about one-third of self-employed individuals and accounts for two-thirds of self-employment income. If we recompute the change in entry rates for this subgroup of self-employed, we find declines that are in line with Decker et al. (2014) and Bayard et al. (2018).

Where we differ with the literature is in our conclusions concerning selection into entrepreneurship. A common finding from previous work is that individuals entering self-employment have lower past labor incomes when compared to peers that are similar but did not enter. As Evans

Consistent with the thesis of Levine and Rubinstein (2018), we find that pooling all self-employed individuals into a single group masks interesting heterogeneity and potentially misleads inference about entrepreneurial activity.
and Leighton (1989) explain, such findings are consistent with sociological views that “misfits,” who are poorer wage earners and more likely to change jobs, are more likely to be self-employed. This is contrary to our findings, which show that most individuals entering self-employment have higher past labor incomes relative to similar peers that did not enter (see Figure 15). Since we also find that there are strong pecuniary motives for many of our switchers and large ex post returns to choosing self-employment, possible top-coding issues with the survey data may be leading to differences in our conclusions.

Another common finding is that individuals that enter self-employment have greater holdings of financial assets. This finding has sparked a large literature emphasizing significant liquidity requirements as impediments to self-employment. See, for example, the work of Evans and Jovanovic (1989), Quadrini (1999), Cagetti and DeNardi (2006), and Buera (2009). There are notable exceptions, namely, Hurst and Lusardi (2004) and Fairlie (2005), who find a limited role for liquid assets as determinants of self-employment. As we showed earlier, when we compare self-employment entrants to comparable non-entrants, we find the latter has higher average asset income. Thus, we view our findings as strengthening the conclusions of Hurst and Lusardi (2004) and Fairlie (2005).

5 Implications for Theory

From our empirical analysis, we find several salient empirical patterns that can be used to inform theories of entrepreneurship. The attached self-employed have persistently high income growth profiles as compared to paid-employed peers. Exit rates are high when individuals are young and decline steeply over the life cycle. The volatility of income changes is higher in self- than paid-employment but does not rise with age. In this section, we analyze a theoretical model that incorporates features motivated by these empirical findings and use it to determine how well the predictions for incomes and growth align with empirical counterparts. Here, we focus on the decision making of young entrepreneurs—those that are attached to their employment status and those that ultimately switch—but the model is sufficiently general to be used in future analyses to study other aspects of the data described earlier.

In the spirit of Bhandari and McGrattan (2021), we incorporate firm-specific investments in self-created intangible assets—customer bases, client lists, inventions, designs, processes—that are needed before production can begin at an optimal scale. In the spirit of Jovanovic (1982), we assume returns on these investments are uncertain because our founders have no previous experience and must learn about their productive capabilities for running a business. As they gain experience, they choose to continue with the business or to discontinue, selling their intangible assets and switching to paid-employment following the exit. When or if they exit depends on the productivity shocks

---

40 See also Alba-Ramirez (1994), Rissman (2003), and Rissman (2007).
these owners face in self-employment and in paid-employment if they do switch.

5.1 Model

Each period, entrepreneurs decide if they will continue running their businesses, sell them, or discontinue without sale and then work for someone else. They condition these decisions on the state \( s \), which depends on financial asset holdings \( a \), business intangible assets \( \kappa \), ability in paid-employment \( \epsilon \), latent ability in self-employment \( z \), and years of experience \( j \). Because \( z \) is latent, the predicted mean \( \mu \) of ability, which depends on past observations of productivity, is also included in the state vector \( s \).

Owners that decide to keep the business choose consumption and inputs for intangible investment and production. Intangible investment requires time, \( h_\kappa \), and expenses, \( e \), which are inputs in the technology \( f_\kappa(h_\kappa, e) \). Production requires the stock of intangible assets, \( \kappa \), hours of the business owner, \( h_y \), and external factors that can be rented, namely, tangible capital, \( k \), and labor \( n \), which are inputs in the technology \( f_y(\kappa, h_y, k, n) \). The prices for the external capital and labor are \( r \) and \( w \), respectively, and taken as given by the businesses. Denoting the value of keeping the business by \( V_k(s) \), we formulate the problem as a dynamic program:

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

and additionally processes for updating \( \epsilon, z \), and \( \mu \) described below. Goods and services sold by the business have a unit price of \( p \), and the capital stocks are assumed to depreciate at rate \( \delta_k \) for tangible capital and \( \delta_\kappa \) for intangible capital. In this problem, the value next period is \( V(s') \) and is the maximum value of the three alternatives: continuing \( V_k(s') \), discontinuing with sale \( V_s(s') \), and discontinuing without sale \( V_w(s') \). The value is a discounted sum of period utilities \( U(c, \ell) \) defined over consumption and leisure, with future values discounted by factor \( \beta \in [0, 1) \).

The entrepreneurial productivity has a permanent unobserved component \( \bar{z}_0 \) plus an idiosyncratic and identically distributed shock \( \eta_j \). Let \( z_j \) denote the value of productivity for an entrepreneur with \( j \) years of experience. In this case,

\[
z_j = \bar{z}_0 + \eta_j,
\]

and we assume that \( \eta_j \) is normally distributed with mean 0 and variance \( \sigma^2_\eta \) that is constant for all \( j \). Let \( \mu_j \) and \( \sigma^2_j \) denote the predicted mean and variance of the latent productivity conditioned on
past signals, that is,

\[ \mu_j = E[\tilde{z}_0 | z_0, z_1, \ldots, z_{j-1}] \]

\[ \sigma^2_j = \text{var}[\tilde{z}_0 | z_0, z_1, \ldots, z_{j-1}] \]

and the distribution of \( z_j \) conditional on the past observations \( z_0, z_1, \ldots, z_{j-1} \) is assumed to be normally distributed with mean \( \mu_j \) and variance \( \sigma^2_j + \sigma^2_\eta \). In this case, the means and variances evolve with one more year of experience as follows:

\[ \mu_{j+1} = \mu_j + \frac{\sigma^2_j}{\sigma^2_j + \sigma^2_\eta} (z_j - \mu_j) \]

\[ \sigma^2_{j+1} = \frac{\sigma^2_j \sigma^2_\eta}{\sigma^2_j + \sigma^2_\eta} \]

The next period value in (13) also depends on the evolution of the paid-employment ability because the entrepreneur can decide to sell or discontinue with business. For tractability, we assume that entrepreneurs that sell or choose to switch to paid-employment do not start a business after that. Relevant to this choice is the evolution of \( \epsilon \) which is Markov with transition probability \( \pi(\epsilon' | \epsilon) \). Then the value of sale is given by

\[ V_s(s) = \max_{c, h, k, n} \left\{ U(c, \ell) + \beta \sum_{\epsilon' | \epsilon} \pi(\epsilon' | \epsilon) V_w(s') \right\} \] (14)

subject to:

\[ a' = (1 + r)a + pe^\delta f_y(\kappa, h_y, k, n) - (r + \delta_k)k - wn + p_\kappa \kappa - c \]

\[ \ell = 1 - h_y \]

with \( a' \geq 0 \) and \( \kappa' = 0 \) and revenues from the sale given by \( p_\kappa \kappa \). The value of working for others, \( V_w \), is standard and given by

\[ V_w(s) = \max_{c, h} \left\{ U(c, \ell) + \beta \sum_{\epsilon' | \epsilon} \pi(\epsilon' | \epsilon) V_w(s') \right\} \] (15)

subject to:

\[ a' = (1 + r)a + wh_y - c \]

\[ \ell = 1 - h_y \]

with \( a' \geq 0 \), where again for tractability, we have assumed that paid employees do not switch to self-employment mid-career.

### 5.2 Quantitative results

Next, we analyze numerical simulations of the entrepreneurial optimization problem and compare predicted growth profiles with empirical counterparts. For our baseline parameterization, we use estimates for preferences and technologies from Bhandari and McGrattan (2021) based on aggregate
data from the BEA’s national accounts and microdata from the Census’s Survey of Business Owners (SBO) and the Pratt’s Stats database of brokered business sales.\textsuperscript{41} We use the IRS microdata to ensure that the variability and persistence of incomes are consistent across model and data.

The functional forms for preferences and technologies used by Bhandari and McGrattan (2021) are given by

\[ U(c, \ell) = (c^{1-\psi} \ell^{\psi})^{1-\sigma}/(1 - \sigma) \]
\[ f_n(h_\kappa, e) = h_\kappa^{\vartheta} e^{1-\vartheta} \]
\[ f_y(\kappa, h_y, k, n) = \kappa^{\phi} k^{\alpha} (\omega h_y^{\rho} + (1 - \omega)n^{\rho})^{\frac{\nu}{\rho}}, \]

with values for parameters listed in Table 6. There are three parameters related to preferences: \( \psi, \sigma, \) and \( \beta. \) Setting the weight on leisure, \( \psi, \) to 58 percent ensures that levels of aggregated business hours are consistent with U.S. totals. The value of 1.5 for \( \sigma \) is standard in the literature. The value of 0.96 for \( \beta \) is consistent with U.S. real returns to capital of roughly 4 percent. In terms of technology parameters, the most relevant for the income and growth profiles is the share of intangible capital in the production of goods and services, \( \phi. \) The share \( \phi \) affects founders’ incentives to invest time and resources in building their business. If the revenue share is small and only external factors are required, then growth of the productive self-employed will be high relative to paid-employed or entrepreneurs that switch out of self-employment early. Bhandari and McGrattan (2021) jointly estimate this parameter, along with two other parameters governing intangible production \( (\vartheta, \rho) \), using information about the intangible share of assets in business sales, the input shares from the BEA input-output tables, and the entry rate of new businesses. They report estimates of \( \phi = 0.15, \vartheta = 0.408, \) and \( \rho = 0.500. \) The remaining production shares, namely, \( \alpha = 0.3, \omega = 0.425, \) and \( \nu = 0.55, \) are based on revenue shares in U.S. private business data. Finally, the depreciation rates used by Bhandari and McGrattan (2021) are based on studies of depreciable and amortizable assets conducted by the BEA and IRS and set equal to \( \delta_k = 4.1 \) percent and \( \delta_\kappa = 5.8 \) percent.

IRS microdata are used to estimate prices and productivities and reported in the lower panels of Table 6. Two prices are pre-set: the interest rate at \( r = 4.1 \) percent, which is consistent with the preference parameters, and the wage rate at \( w = 1, \) which is a normalization. Given this value of the wage, we set the price of goods and services \( p \) equal to 1.5 to ensure that the relative income for the young entrepreneurs—stayers versus switchers—is consistent with IRS data at age 40. The price per unit of self-created capital, \( p_\kappa, \) if sold is set equal to 1.6 and chosen to generate ratios of the business value to seller’s wage bill between 2 and 3, consistent with U.S. private business sales. For entrepreneurial productivities, we normalize the predicted mean \( \mu_0 \) to 0 (along with

\textsuperscript{41}Estimating the full model requires linking all business filings for the self-employed and is beyond the scope of this paper.
the average $\bar{z}_0$). Shock variances driving entrepreneurial income shocks are chosen to ensure that the 90–10 differences and autocorrelations in business income changes are the same in the model and the IRS data. This implies values for the predicted initial productivity variance, $\sigma^2_0$, equal to 0.5 percent and the idiosyncratic variance, $\sigma^2_\eta$, equal to 0.4 percent. Similarly, the parameters governing paid-employment income are set to ensure that the 90–10 differences and autocorrelations of employee income changes—relevant for those exiting to paid-employment—are the same in the model and the IRS data. To model $\pi(\epsilon'|\epsilon)$, we use the method of Tauchen (1986) to approximate a continuous autoregressive process as a Markov chain. When simulating our data, we use 11 states, a persistence parameter of 0.7 and a standard deviation of 0.1.

Given parameter estimates, we now use our laboratory to simulate income and growth profiles for a large sample of entrepreneurs. In the simulations, we assume a full life cycle of 60 years but report results for ages 25 to 40 in order to compare our predictions to the data on young entrepreneurs—both the stayers and switchers. We assume that these young start-ups have no assets and little in the way of transfers—just 0.01 so that initial consumption is not zero. As we vary these choices, we find that there is only a small impact on production decisions for entrepreneurs with high predicted productivity levels, who want to scale up as quickly as possible.

To make the model and data results comparable, we use the counts by employment status from the 1975 cohort, which includes individuals between 25 and 40 during our sample period. For example, we know how many are self-employed at age 25, how many at age 26, and so on. Using these counts, we find a roughly constant entry rate into self-employment between ages 25 and 31, with rates on the order of 11 percent per year. Using this constant rate, we extrapolate back to age 22, which is before we see them in the sample. To compute theoretical predictions, we simulate data for 22-year-olds, 23-year-olds, and so on, and then use the actual counts of self-, paid-, and non-employed to weight the model-generated incomes (which are equal to the average wage for paid-employed and zero for non-employed). Then, we construct income and growth profiles for two groups from the model simulations: self-employed stayers and self-employed switchers. Both groups have at least five years of either self- or paid-employment experience prior to age 35. After 35, the self-employed stayers have continued on in self-employment, and the self-employed switchers have discontinued or sold their business and switched to paid-employment. Weights from the 1975 cohort are then used to add up the stayers and the switchers at different ages.

In Figure 20, we show the differenced income growth from the data (that is, differences in the two profiles in Figure 17 against the predicted profile). We should note that we have not included any economy-wide technological changes or alternative sources of growth that would be common to individuals with different employment status. In fact, in the model, the outside opportunity of paid employment is a flat income profile when averaged. Therefore, we compute differences in growth profiles for both the model and data so they are comparable. Both show a hump-shaped growth
profile. In the model, we generate this because of two key features: learning and investment. If we abstract from either, then we are unable to generate this pattern.

Consider first the role of learning. At the start of the simulations, the self-employed that ultimately stay or switch look the same. They all start with an initial prediction \( \mu_0 \) of 0 and the same variance on the productivity signal. After that, they gain experience, and those that ultimately exit self-employment have a mean prediction for their productivity that has fallen over time. Importantly, the fall in the mean prediction leads these entrepreneurs to reduce their investments in self-created intangibles over time. Less investment means less growth in subsequent years and, thus, an eventual exit to better opportunities in paid-employment. If there is greater certainty about the entrepreneurs’ productive capability—that is, if the variance \( \sigma^2 \eta \) is lowered—exits occur earlier. If it is sufficiently low, then exits occur immediately, and we would not observe any entrepreneurs waiting five years before switching to paid-employment.

The second key feature is the investment in \( \kappa \) made by entrepreneurs. For our baseline parameters, we find that roughly 10 percent of available time is used initially to build \( \kappa \). By age 35, investment is close to zero for the switchers but around 10 percent for the stayers. Noteworthy is the fact that entrepreneurs who ultimately stay in self-employment start increasing their investments immediately in order to quickly build up their intangible capital stocks. These investments ultimately pay off in higher incomes later. As they build the intangible stock, entrepreneurs start to substitute external hours from paid employees for their own time in production of goods and services. For the entrepreneurs that continue past the age of 35, we find a steady drop in own hours of production and a scaling up of the business as they hire external labor and capital. By age 40, the ratio of external to internal hours is roughly 6 times. For switchers, we find almost no scaling up.

The impact that investment has on growth depends importantly on the revenue share for the self-created intangibles, \( \kappa \). In Figure 21, we report the predicted growth differential estimates as we vary \( \phi \) and thus the revenue share. For the simulations, we hold all other parameters and prices fixed and thus find similar estimates for the incomes of our two groups at age 40, even though the life-cycle growth patterns differ. As Panel A of Figure 21 shows, the choice of \( \phi \) can have a large effect on the differential growth between entrepreneurs that continue and those that exit. In the baseline parameterization, we set \( \phi \) to 0.15. In the figure, we show growth differentials as we lower \( \phi \) to 0.1 and even further to 0.05. The associated investments in each case are shown in Panel B of Figure 21. With a larger revenue share for intangibles, the owner is incentivized to invest and the growth in income slower and more persistent. When the share is lowered, investments decline more quickly, and the growth in incomes occurs earlier. In this case, the owners rely more on external factors and scale up the business at an earlier age. When and how much they scale up depends on the specification of hours in production, for example, the share of owner time,
\( \omega \), and its substitutability with external labor, \( \rho \). If owner time and employee time are highly substitutable, then the owner can create the intangible asset—say, the list of clients—and the employees can work with them. Relatedly, the external resource requirement versus own time in intangible capital production governed by \( \vartheta \) is relevant for time use early in the career. But varying these parameters does not change the overall message that our predictions for the growth differentials depend importantly on incorporating nontrivial firm-specific investments.

Overall, we find that the model does surprisingly well in generating growth differentials that are consistent with the young entrepreneurs in the IRS sample.

6 Conclusions

Much has been written about the nature of entrepreneurship, but our knowledge base is built up from analyses of very different samples of individuals, which on the whole provide a narrative reminiscent of the parable of the blind men and an elephant. Each man learns about the elephant by touching only one part of the body, drawing conclusions that the elephant is like a wall, snake, spear, tree, fan, or rope, depending on what part they had touched. Analogously, the literature on entrepreneurship has an array of narratives, describing the typical business owner in many possible ways: as a gig worker seeking flexible arrangements, a misfit avoiding unemployment spells, an inventor seeking venture capital, a wealthy individual with no financing needs, or a tax dodger. To provide a more complete picture of the nature of entrepreneurship, we used U.S. administrative tax data to assemble a novel longitudinal database of business owners—one that is suitable for analyzing patterns of income growth and determinants of entrepreneurial choice for a large population of self-employed individuals.

Critical to the analysis was our notion of employment attachment, whether individuals were in the same employment status—say, self-, paid-, or non-employment—for most of the sample or were mostly switching. Comparisons of income and growth profiles for the attached self- and paid-employed revealed a striking contrast: the average income growth profiles of the self-employed are much higher and more hump-shaped than those for paid-employed peers with the same characteristics. Comparisons of income changes for these groups revealed that dispersion in incomes declines over the life cycle regardless of employment status. Analysis of the switching into and out of self-employment also yielded new insights relative to earlier work. We found that individuals entering self-employment have higher past wage income and lower past asset income than peers not entering, which is contrary to earlier findings based on survey data.

We hope and expect that the empirical results of the paper will motivate new theories of entrepreneurship, which can be used to provide better tools for tax administrators and policymakers. In our view, critical inputs will include firm-specific investments, incomplete information about
entrepreneurial productivity, and other mechanisms that imply a slow adjustment to the optimal size of operation.
References


Table 1: Main Sample Summary Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Total Sample</th>
<th>Attached PE</th>
<th>Attached SE</th>
<th>Almost PE</th>
<th>Almost SE</th>
<th>Mostly Switching PE</th>
<th>Mostly Switching SE</th>
<th>Any NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals (Mil.)</td>
<td>80.0</td>
<td>41.7</td>
<td>2.0</td>
<td>0.5</td>
<td>0.3</td>
<td>2.4</td>
<td>33.0</td>
<td></td>
</tr>
<tr>
<td>Shares (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counts</td>
<td>100</td>
<td>52.2</td>
<td>2.4</td>
<td>0.7</td>
<td>0.3</td>
<td>3.0</td>
<td>41.3</td>
<td></td>
</tr>
<tr>
<td>Total income</td>
<td>100</td>
<td>66.5</td>
<td>7.6</td>
<td>1.2</td>
<td>1.0</td>
<td>6.1</td>
<td>17.6</td>
<td></td>
</tr>
<tr>
<td>PE income</td>
<td>100</td>
<td>75.9</td>
<td>0.9</td>
<td>1.3</td>
<td>0.4</td>
<td>4.3</td>
<td>17.2</td>
<td></td>
</tr>
<tr>
<td>SE income</td>
<td>100</td>
<td>4.2</td>
<td>51.5</td>
<td>0.8</td>
<td>5.3</td>
<td>18.5</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>Incomes (2012$, Thous.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean, PE plus SE income</td>
<td>49.0</td>
<td>62.5</td>
<td>152.1</td>
<td>89.5</td>
<td>146.9</td>
<td>98.7</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>Percentiles, 10th</td>
<td>6.0</td>
<td>22.7</td>
<td>17.0</td>
<td>21.5</td>
<td>21.1</td>
<td>17.0</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>25th</td>
<td>15.7</td>
<td>32.3</td>
<td>30.8</td>
<td>32.4</td>
<td>38.3</td>
<td>28.0</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>50th</td>
<td>32.9</td>
<td>47.5</td>
<td>65.7</td>
<td>52.2</td>
<td>72.8</td>
<td>50.6</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>75th</td>
<td>56.9</td>
<td>70.5</td>
<td>151.1</td>
<td>89.3</td>
<td>146.3</td>
<td>99.0</td>
<td>25.4</td>
<td></td>
</tr>
<tr>
<td>90th</td>
<td>92.4</td>
<td>105.6</td>
<td>330.9</td>
<td>157.1</td>
<td>303.3</td>
<td>197.0</td>
<td>42.0</td>
<td></td>
</tr>
<tr>
<td>Mean, PE income</td>
<td>42.6</td>
<td>61.9</td>
<td>16.2</td>
<td>81.8</td>
<td>45.9</td>
<td>59.4</td>
<td>17.8</td>
<td></td>
</tr>
<tr>
<td>Percentiles, 10th</td>
<td>3.3</td>
<td>22.6</td>
<td>0.0</td>
<td>20.4</td>
<td>4.5</td>
<td>10.3</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>25th</td>
<td>12.8</td>
<td>32.3</td>
<td>0.0</td>
<td>30.8</td>
<td>8.2</td>
<td>17.7</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>50th</td>
<td>30.3</td>
<td>47.4</td>
<td>2.0</td>
<td>48.6</td>
<td>17.5</td>
<td>31.7</td>
<td>11.8</td>
<td></td>
</tr>
<tr>
<td>75th</td>
<td>53.4</td>
<td>70.1</td>
<td>10.4</td>
<td>80.7</td>
<td>41.4</td>
<td>58.7</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td>90th</td>
<td>84.4</td>
<td>104.6</td>
<td>32.6</td>
<td>139.6</td>
<td>91.0</td>
<td>108.9</td>
<td>38.1</td>
<td></td>
</tr>
<tr>
<td>Mean, SE income</td>
<td>6.5</td>
<td>0.5</td>
<td>135.9</td>
<td>7.7</td>
<td>101.0</td>
<td>39.3</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>Percentiles, 10th</td>
<td>-0.0</td>
<td>-0.1</td>
<td>14.5</td>
<td>-4.5</td>
<td>10.0</td>
<td>0.4</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>25th</td>
<td>0.0</td>
<td>0.0</td>
<td>26.5</td>
<td>0.1</td>
<td>20.4</td>
<td>5.6</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>50th</td>
<td>0.0</td>
<td>0.0</td>
<td>57.3</td>
<td>3.1</td>
<td>45.7</td>
<td>14.3</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>75th</td>
<td>0.4</td>
<td>0.0</td>
<td>134.2</td>
<td>8.8</td>
<td>101.3</td>
<td>37.5</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>90th</td>
<td>6.8</td>
<td>0.8</td>
<td>297.8</td>
<td>21.0</td>
<td>218.0</td>
<td>90.0</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Education and skills (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College-educated</td>
<td>52.8</td>
<td>60.9</td>
<td>63.4</td>
<td>70.3</td>
<td>74.7</td>
<td>67.3</td>
<td>40.5</td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>52.4</td>
<td>58.2</td>
<td>60.8</td>
<td>65.3</td>
<td>66.1</td>
<td>63.6</td>
<td>43.6</td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td>58.6</td>
<td>66.5</td>
<td>61.2</td>
<td>69.6</td>
<td>71.4</td>
<td>66.4</td>
<td>47.7</td>
<td></td>
</tr>
<tr>
<td>Manual</td>
<td>37.6</td>
<td>35.3</td>
<td>35.3</td>
<td>35.0</td>
<td>28.0</td>
<td>34.3</td>
<td>41.0</td>
<td></td>
</tr>
<tr>
<td>Primary industry shares (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>1.1</td>
<td>0.9</td>
<td>1.6</td>
<td>1.3</td>
<td>1.6</td>
<td>1.2</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Mining</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.5</td>
<td>0.6</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

See notes at end of table.
Table 1: Main Sample Summary Statistics (cont.)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Total Sample</th>
<th>Attached PE</th>
<th>Attached SE</th>
<th>Almost PE</th>
<th>Almost SE</th>
<th>Mostly Switching PE</th>
<th>Mostly Switching SE</th>
<th>Any NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary industry shares (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>7.2</td>
<td>5.3</td>
<td>17.2</td>
<td>10.3</td>
<td>15.0</td>
<td>12.7</td>
<td>8.6</td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>11.5</td>
<td>14.3</td>
<td>3.6</td>
<td>10.5</td>
<td>4.7</td>
<td>6.3</td>
<td>8.9</td>
<td></td>
</tr>
<tr>
<td>Wholesale trade</td>
<td>3.6</td>
<td>4.0</td>
<td>3.4</td>
<td>4.2</td>
<td>4.0</td>
<td>3.8</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Retail trade</td>
<td>8.9</td>
<td>8.1</td>
<td>7.9</td>
<td>8.4</td>
<td>9.5</td>
<td>8.4</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Transportation</td>
<td>3.4</td>
<td>3.1</td>
<td>4.9</td>
<td>4.8</td>
<td>4.1</td>
<td>5.4</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>1.6</td>
<td>1.8</td>
<td>0.9</td>
<td>2.5</td>
<td>1.3</td>
<td>1.6</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Finance</td>
<td>3.3</td>
<td>3.7</td>
<td>4.5</td>
<td>4.3</td>
<td>4.4</td>
<td>4.2</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>Real estate</td>
<td>2.3</td>
<td>1.8</td>
<td>4.7</td>
<td>3.4</td>
<td>4.8</td>
<td>4.0</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>Professional services</td>
<td>9.1</td>
<td>8.8</td>
<td>16.0</td>
<td>15.3</td>
<td>17.2</td>
<td>16.6</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>0.7</td>
<td>0.9</td>
<td>0.1</td>
<td>0.6</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>4.9</td>
<td>3.8</td>
<td>4.4</td>
<td>4.6</td>
<td>3.9</td>
<td>4.4</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
<td>0.7</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Health care</td>
<td>6.4</td>
<td>5.5</td>
<td>10.6</td>
<td>6.4</td>
<td>10.3</td>
<td>9.3</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>Arts</td>
<td>1.0</td>
<td>0.8</td>
<td>1.8</td>
<td>1.5</td>
<td>2.1</td>
<td>1.6</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Accommodation</td>
<td>4.4</td>
<td>3.3</td>
<td>3.4</td>
<td>4.0</td>
<td>4.8</td>
<td>4.2</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>Other services</td>
<td>3.8</td>
<td>2.1</td>
<td>11.1</td>
<td>4.2</td>
<td>9.0</td>
<td>7.7</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>Other NAICS</td>
<td>13.7</td>
<td>17.4</td>
<td>0.6</td>
<td>11.3</td>
<td>0.6</td>
<td>5.0</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>Missing NAICS</td>
<td>11.8</td>
<td>13.3</td>
<td>2.2</td>
<td>0.9</td>
<td>1.5</td>
<td>1.9</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>50.2</td>
<td>52.8</td>
<td>82.3</td>
<td>74.6</td>
<td>79.9</td>
<td>74.6</td>
<td>42.5</td>
<td></td>
</tr>
<tr>
<td>Mostly married (%)</td>
<td>63.8</td>
<td>68.1</td>
<td>78.7</td>
<td>72.4</td>
<td>81.8</td>
<td>73.7</td>
<td>56.4</td>
<td></td>
</tr>
<tr>
<td>Has children (%)</td>
<td>82.1</td>
<td>82.6</td>
<td>85.0</td>
<td>86.2</td>
<td>87.1</td>
<td>85.9</td>
<td>81.0</td>
<td></td>
</tr>
<tr>
<td>Number of children (mean)</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.5</td>
<td>2.3</td>
<td>2.4</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Other Incomes (2012$, Thous.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spousal wages, Mean</td>
<td>32.8</td>
<td>30.9</td>
<td>25.9</td>
<td>25.5</td>
<td>29.9</td>
<td>27.7</td>
<td>36.1</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>17.9</td>
<td>20.7</td>
<td>8.6</td>
<td>10.6</td>
<td>13.4</td>
<td>11.3</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>Asset income, Mean</td>
<td>8.7</td>
<td>4.6</td>
<td>57.7</td>
<td>17.4</td>
<td>62.8</td>
<td>25.9</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0</td>
<td>0.1</td>
<td>1.9</td>
<td>0.1</td>
<td>2.6</td>
<td>0.3</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>UI income, Mean</td>
<td>0.4</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Notes: PE=paid-employed, SE=self-employed, and NE=non-employed. See Sections 3 and 4 for details on the samples and subgroups. To ensure that no confidential information is disclosed, reported percentiles are computed as an average of observations around the value listed in the table.
Table 2: Largest Contributors to Attached Employee Growth Gap

<table>
<thead>
<tr>
<th>Cumulative Share</th>
<th>NAICS</th>
<th>Cohort</th>
<th>Cognitive</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.6</td>
<td>62</td>
<td>1960s</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>18.8</td>
<td>54</td>
<td>1960s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.5</td>
<td>52</td>
<td>1960s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.7</td>
<td>62</td>
<td>1970s</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>31.5</td>
<td>54</td>
<td>1960s</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>34.3</td>
<td>44</td>
<td>1960s</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>36.9</td>
<td>62</td>
<td>1970s</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>39.5</td>
<td>54</td>
<td>1970s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.0</td>
<td>52</td>
<td>1970s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.6</td>
<td>62</td>
<td>1960s</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>47.1</td>
<td>23</td>
<td>1960s</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>49.4</td>
<td>23</td>
<td>1970s</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>51.7</td>
<td>54</td>
<td>1970s</td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The sample underlying this table includes all attached self- and paid-employed subgroups. Shares of the contribution to the self- and paid-employment growth gap—which is displayed in Panel B of Figure 2—are computed for ages between 30 and 39 and then cumulated. Results are reported for the top groups contributing at least 50 percent to the gap. All top contributors are male, married, college-educated, interpersonally skilled, and have kids.
Table 3: Transition Probabilities for Incomes in Levels

### Attached Self-Employed

<table>
<thead>
<tr>
<th>Income, Period $t$</th>
<th>$&lt;0$</th>
<th>0-7</th>
<th>7-14</th>
<th>14-27</th>
<th>27-54</th>
<th>54-106</th>
<th>106-208</th>
<th>208-410</th>
<th>410-806</th>
<th>806-1587</th>
<th>$&gt;1587$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income, Period $t-1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-7</td>
<td>6</td>
<td>19</td>
<td>33</td>
<td>22</td>
<td>12</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7-14</td>
<td>2</td>
<td>5</td>
<td>49</td>
<td>31</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-27</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>57</td>
<td>22</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27-54</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>16</td>
<td>57</td>
<td>19</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54-106</td>
<td>2</td>
<td></td>
<td>1</td>
<td>3</td>
<td>17</td>
<td>58</td>
<td>16</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106-208</td>
<td>2</td>
<td></td>
<td>1</td>
<td>3</td>
<td>19</td>
<td>59</td>
<td>14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208-410</td>
<td>2</td>
<td></td>
<td>1</td>
<td>3</td>
<td>19</td>
<td>60</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>410-806</td>
<td>2</td>
<td></td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>20</td>
<td>59</td>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>806-1587</td>
<td>3</td>
<td></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>22</td>
<td>55</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&gt;1587$</td>
<td>4</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distribution: 3 1 6 15 20 20 10 5 2 1

### Attached Paid-Employed

<table>
<thead>
<tr>
<th>Income, Period $t$</th>
<th>$&lt;0$</th>
<th>0-7</th>
<th>7-14</th>
<th>14-27</th>
<th>27-54</th>
<th>54-106</th>
<th>106-208</th>
<th>208-410</th>
<th>410-806</th>
<th>806-1587</th>
<th>$&gt;1587$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income, Period $t-1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-7</td>
<td>29</td>
<td>7</td>
<td>11</td>
<td>19</td>
<td>20</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-14</td>
<td>1</td>
<td>19</td>
<td>33</td>
<td>31</td>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-27</td>
<td>5</td>
<td>46</td>
<td>36</td>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27-54</td>
<td>1</td>
<td>7</td>
<td>81</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54-106</td>
<td>1</td>
<td>9</td>
<td>84</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106-208</td>
<td>1</td>
<td>14</td>
<td>78</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208-410</td>
<td>2</td>
<td>20</td>
<td>67</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>410-806</td>
<td>1</td>
<td>5</td>
<td>24</td>
<td>59</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>806-1587</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>25</td>
<td>48</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&gt;1587$</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>21</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distribution: 0 1 3 14 39 32 9 2 0 0 0

Notes: The sample underlying these transition matrices includes all attached self- and paid-employed individuals. Incomes are in thousands of 2012 dollars. Element $(i, j)$ of each matrix is the probability of having income in bin $i$ in tax year $t-1$ and income in bin $j$ in tax year $t$. Elements with values below 0.5 percent are left blank. The stationary distribution is listed in the last lines of each matrix.
Table 4: Transition Probabilities for Residual Growth Rates

**Attached Self-Employed**

<table>
<thead>
<tr>
<th>% Change, Period t</th>
<th>&lt; 628</th>
<th>-628 - 185</th>
<th>-185 - 54</th>
<th>-54 - 16</th>
<th>-16 - 5</th>
<th>-5</th>
<th>-5</th>
<th>16</th>
<th>54 - 185</th>
<th>185 - 628</th>
<th>&gt; 628</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 628</td>
<td>3 6 7 4 2 2 2 9 36 18 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-628 - 185</td>
<td>2 8 11 5 2 2 2 8 30 25 6</td>
<td>1 4 13 11 4 4 4 13 24 16 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-185 - 54</td>
<td>1 4 13 11 4 4 4 13 24 16 5</td>
<td>1 10 21 10 9 10 21 15 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-54 - 16</td>
<td>1 10 21 10 9 10 21 15 3</td>
<td>7 22 15 16 14 18 7 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-16 - 5</td>
<td>7 22 15 16 14 18 7 1</td>
<td>6 22 18 17 14 16 6 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5 - 5</td>
<td>6 22 18 17 14 16 6 1</td>
<td>5 22 17 16 13 16 6 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5 - 16</td>
<td>5 22 17 16 13 16 6 1</td>
<td>4 22 17 16 13 16 6 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 - 54</td>
<td>4 22 17 16 13 16 6 1</td>
<td>1 9 27 14 12 11 16 8 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 - 185</td>
<td>1 2 16 27 10 8 8 14 10 3 1</td>
<td>1 23 22 7 6 6 13 12 5 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>185 - 628</td>
<td>1 23 22 7 6 6 13 12 5 1</td>
<td>2 5 28 20 6 5 5 11 11 5 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 628</td>
<td>2 5 28 20 6 5 5 11 11 5 1</td>
<td>2 5 28 20 6 5 5 11 11 5 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Attached Paid-Employed**

<table>
<thead>
<tr>
<th>% Change, Period t</th>
<th>&lt; 628</th>
<th>-628 - 185</th>
<th>-185 - 54</th>
<th>-54 - 16</th>
<th>-16 - 5</th>
<th>-5</th>
<th>-5</th>
<th>16</th>
<th>54 - 185</th>
<th>185 - 628</th>
<th>&gt; 628</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 628</td>
<td>1 2 4 3 2 2 2 8 51 18 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-628 - 185</td>
<td>1 3 5 4 1 1 2 8 39 33 5</td>
<td>1 3 5 4 1 1 2 8 39 33 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-185 - 54</td>
<td>1 3 5 4 1 1 2 8 39 33 5</td>
<td>5 10 5 5 5 14 34 19 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-54 - 16</td>
<td>5 10 5 5 5 14 34 19 3</td>
<td>4 16 15 17 15 21 12 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-16 - 5</td>
<td>4 16 15 17 15 21 12 1</td>
<td>2 10 34 23 18 10 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5 - 5</td>
<td>2 10 34 23 18 10 2</td>
<td>1 8 45 20 18 7 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5 - 16</td>
<td>1 8 45 20 18 7 1</td>
<td>2 11 35 23 19 9 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 - 54</td>
<td>2 11 35 23 19 9 1</td>
<td>3 20 24 21 16 13 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 - 185</td>
<td>3 20 24 21 16 13 3</td>
<td>6 23 15 19 14 15 6 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>185 - 628</td>
<td>6 23 15 19 14 15 6 1</td>
<td>11 18 13 17 14 18 8 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 628</td>
<td>11 18 13 17 14 18 8 1</td>
<td>17 17 11 14 12 18 9 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The sample underlying these transition matrices includes all attached self- and paid-employed individuals. Element \((i, j)\) of each matrix is the probability of having the residual growth rate \((\Delta \xi\text{ divided by lagged income})\) in bin \(i\) in tax year \(t - 1\) and the residual growth rate in bin \(j\) in tax year \(t\). Elements with values below 0.5 percent are left blank. The stationary distribution is listed in the last lines of each matrix.
Table 5: Income Shares Held by Each Employment Group

### Total Income

<table>
<thead>
<tr>
<th>Percentile Group</th>
<th>Total Sample</th>
<th>Attached</th>
<th>Almost</th>
<th>Mostly</th>
<th>Any</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10th</td>
<td>-0.3</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>10th – 25th</td>
<td>4.1</td>
<td>1.2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>25th – 75th</td>
<td>35.5</td>
<td>24.2</td>
<td>0.9</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>75th – 90th</td>
<td>21.3</td>
<td>16.7</td>
<td>1.1</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>&gt; 90th</td>
<td>39</td>
<td>23.1</td>
<td>6.3</td>
<td>0.7</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Paid-Employment Income

<table>
<thead>
<tr>
<th>Percentile Group</th>
<th>Total Sample</th>
<th>Attached</th>
<th>Almost</th>
<th>Mostly</th>
<th>Any</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10th</td>
<td>1.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>10th – 25th</td>
<td>4.3</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>25th – 75th</td>
<td>38.5</td>
<td>28.0</td>
<td>0.1</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>75th – 90th</td>
<td>22.4</td>
<td>19.3</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>&gt; 90th</td>
<td>33.7</td>
<td>26.2</td>
<td>0.9</td>
<td>0.7</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Self-Employment Income

<table>
<thead>
<tr>
<th>Percentile Group</th>
<th>Total Sample</th>
<th>Attached</th>
<th>Almost</th>
<th>Mostly</th>
<th>Any</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10th</td>
<td>-8.7</td>
<td>-1.7</td>
<td>-1.7</td>
<td>-0.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>10th – 25th</td>
<td>2.6</td>
<td>-0.0</td>
<td>0.4</td>
<td>-0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>25th – 75th</td>
<td>16.7</td>
<td>0.7</td>
<td>5.8</td>
<td>0.1</td>
<td>0.6</td>
</tr>
<tr>
<td>75th – 90th</td>
<td>14.4</td>
<td>1.0</td>
<td>6.8</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>&gt; 90th</td>
<td>75.1</td>
<td>3.7</td>
<td>40.0</td>
<td>0.6</td>
<td>5.3</td>
</tr>
</tbody>
</table>

**Notes:** The sample underlying this table includes all individuals in the baseline sample summarized in the first column of Table 1. The income shares are total income attributed to the group listed in the column headings divided by the total in the ranked group.
Table 6: Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Expression</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preferences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure weight</td>
<td>$\psi$</td>
<td>0.580</td>
</tr>
<tr>
<td>Intertemporal elasticity inverse</td>
<td>$\sigma$</td>
<td>1.500</td>
</tr>
<tr>
<td>Discount factor</td>
<td>$\beta$</td>
<td>0.960</td>
</tr>
<tr>
<td><strong>Technologies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owner hours share, intangible production</td>
<td>$\vartheta$</td>
<td>0.408</td>
</tr>
<tr>
<td>Hours substitution parameter, goods production</td>
<td>$\rho$</td>
<td>0.500</td>
</tr>
<tr>
<td>Intangible capital share, goods production</td>
<td>$\phi$</td>
<td>0.150</td>
</tr>
<tr>
<td>Fixed asset share, goods production</td>
<td>$\alpha$</td>
<td>0.300</td>
</tr>
<tr>
<td>Owner hours share, goods production</td>
<td>$\omega$</td>
<td>0.425</td>
</tr>
<tr>
<td>Intangible capital depreciation</td>
<td>$\delta$</td>
<td>0.058</td>
</tr>
<tr>
<td>Fixed asset depreciation</td>
<td>$\delta_k$</td>
<td>0.041</td>
</tr>
<tr>
<td><strong>Prices</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest rate</td>
<td>$r$</td>
<td>0.041</td>
</tr>
<tr>
<td>Hired labor</td>
<td>$w$</td>
<td>1.000</td>
</tr>
<tr>
<td>Goods and services</td>
<td>$p$</td>
<td>1.500</td>
</tr>
<tr>
<td>Intangible capital</td>
<td>$p_k$</td>
<td>1.600</td>
</tr>
<tr>
<td><strong>Entrepreneurial productivity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial predicted mean</td>
<td>$\mu_0$</td>
<td>0.000</td>
</tr>
<tr>
<td>Initial predicted variance</td>
<td>$\sigma_{\mu}^2$</td>
<td>0.005</td>
</tr>
<tr>
<td>Idiosyncratic shock variance</td>
<td>$\sigma_{\eta}^2$</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Paid-employment shocks</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persistence</td>
<td>$\rho_\epsilon$</td>
<td>0.700</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$\sigma_\epsilon$</td>
<td>0.100</td>
</tr>
</tbody>
</table>
Note: The sample underlying these figures includes all attached self- and paid-employed subgroups. The figure reports weighted averages of the estimated time effects for groups $g$ at time $t$, that is, $\Delta \beta_{g,t}$, which is divided by average income for group $g$ in year $t$, $\overline{y}_{g,t}$. Weights are constructed from group counts.
Figure 2: Income and Growth Profiles

A. Income Profiles

![Graph showing income profiles for self-employed and paid-employed individuals over age]

B. Growth Profiles

![Graph showing growth profiles for self-employed and paid-employed individuals over age]

Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups. Panel A shows the weighted averages of integrated incomes $Y_g(a)$ for subgroup $g$ at age $a$. Panel B shows weighted averages of the associated growth by age, $Y_g(a) - Y_g(a-1)$. In both panels, weights are constructed from sample counts, $N_g^a / \sum_g N_g^a$. 

50
Figure 3: Growth Profiles by Gender

A. Men

Age  | Income Growth (thousands of 2012 dollars)  
--- | ---  
25  | SE incomes: Y(25) = 44,608  
55  | Y(55) = 225,681  
    | PE incomes: Y(25) = 36,498  
    | Y(55) = 102,523  

B. Women

Age  | Income Growth (thousands of 2012 dollars)  
--- | ---  
25  | SE incomes: Y(25) = 31,603  
55  | Y(55) = 136,818  
    | PE incomes: Y(25) = 30,363  
    | Y(55) = 72,354  

Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups, which are further subdivided into those for men (Panel A) and women (Panel B). Both panels show weighted averages of subgroup g growth for each age a, Y_g(a) - Y_g(a-1), with weights constructed from sample counts, N_g / \sum N_g.
Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups, which are further subdivided into those for the married most years (Panel A) and those that are not (Panel B). Both panels show weighted averages of subgroup $g$ growth for each age $a$, $Y_g(a) - Y_g(a-1)$, with weights constructed from sample counts, $N_g^a / \sum_g N_g^a$. 
Figure 5: Growth Profiles by Education

A. College-educated

Income Growth (thousands of 2012 dollars)

Self-employed
Paid-employed

SE incomes:
\[Y(25) = 54,664\]
\[Y(55) = 320,656\]

PE incomes:
\[Y(25) = 36,114\]
\[Y(55) = 111,363\]

Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups, which are further subdivided into those classified as college-educated (Panel A) and those that are not (Panel B). Both panels show weighted averages of subgroup \(g\) growth for each age \(a\), \(Y_g(a) - Y_g(a-1)\), with weights constructed from sample counts, \(N_g^a / \sum_g N_g^a\).
Figure 6: Growth Profiles by Cognitive Skill

A. Cognitively Skilled

B. Not Cognitively Skilled

Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups, which are further subdivided into those classified as cognitively skilled (Panel A) and those that are not (Panel B). Both panels show weighted averages of subgroup $g$ growth for each age $a$, $Y_g(a) - Y_g(a-1)$, with weights constructed from sample counts, $N_g^a / \sum_g N_g^a$. 

54
Figure 7: Growth Profiles by Interpersonal Skill

A. Interpersonally Skilled

![Graph showing growth profiles for interpersonally skilled individuals.](image)

- SE incomes:
  - $Y_{(25)} = 52,565$
  - $Y_{(55)} = 306,862$

- PE incomes:
  - $Y_{(25)} = 35,141$
  - $Y_{(55)} = 104,617$

B. Not Interpersonally Skilled

![Graph showing growth profiles for not interpersonally skilled individuals.](image)

- SE incomes:
  - $Y_{(25)} = 33,043$
  - $Y_{(55)} = 92,258$

- PE incomes:
  - $Y_{(25)} = 30,548$
  - $Y_{(55)} = 55,296$

Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups, which are further subdivided into those classified as interpersonally skilled (Panel A) and those that are not (Panel B). Both panels show weighted averages of subgroup $g$ growth for each age $a$, $Y_g(a) - Y_g(a - 1)$, with weights constructed from sample counts, $N_g / \sum_g N_g^a$. 

55
Figure 8: Growth Profiles by Manual Skill

A. Manually Skilled

<table>
<thead>
<tr>
<th>Age</th>
<th>SE incomes:</th>
<th>PE incomes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>$Y_{25} = 31,716$</td>
<td>$Y_{25} = 31,842$</td>
</tr>
<tr>
<td>55</td>
<td>$Y_{55} = 102,806$</td>
<td>$Y_{55} = 59,877$</td>
</tr>
</tbody>
</table>

B. Not Manually Skilled

<table>
<thead>
<tr>
<th>Age</th>
<th>SE incomes:</th>
<th>PE incomes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>$Y_{25} = 50,741$</td>
<td>$Y_{25} = 34,852$</td>
</tr>
<tr>
<td>55</td>
<td>$Y_{55} = 280,154$</td>
<td>$Y_{55} = 105,086$</td>
</tr>
</tbody>
</table>

Notes: The sample underlying these figures includes all attached self- and paid-employed subgroups, which are further subdivided into those classified as manually skilled (Panel A) and those that are not (Panel B). Both panels show weighted averages of subgroup growth for each age $a$, $Y_g(a) - Y_g(a - 1)$, with weights constructed from sample counts, $N_g / \sum_g N_g$. 

56
A. Professional Services

B. Health Care

Notes: The sample underlying these figures includes individuals in the attached self- and paid-employed subgroups that work in professional services (Panel A) and health care (Panel B). Both panels show weighted averages of subgroup $g$ growth for each age $a$, $Y_g(a) - Y_g(a-1)$, with weights constructed from sample counts, $N_g / \sum_g N_g$.
Figure 10: Growth Profiles of Mostly Switchers and Any Non-Employed

A. Mostly Switchers versus Self-Employed

B. Any Non-employed versus Self-Employed

Notes: The sample underlying these figures includes individuals in the mostly switchers and any-non-employment subgroups. Both panels show weighted averages of subgroup $g$ growth for each age $a$, $Y_g(a) - Y_g(a-1)$, with weights constructed from sample counts, $N^a_g / \sum_g N^a_g$. Results are compared to the growth of the attached self-employed from Figure 2.
Figure 11: Age-over-Age Growth in Incomes

A. Attached Self-Employed

B. Attached Paid-Employed

Notes: The sample underlying these figures includes all attached self- and paid-employed individuals. For each individual, we compute the age-over-age percentage change in income, $\Delta y_{i,a}/|y_{i,a-1}|$, and plot selected percentiles of these changes for the self-employed (Panel A) and paid-employed (Panel B).
Figure 12: Age-over-Age Growth in Regression Residuals

A. Attached Self-Employed

B. Attached Paid-Employed

Notes: The sample underlying these figures includes all attached self- and paid-employed individuals. For each individual, we compute the age-over-age change in the residual normalized by previous year income, $\Delta \epsilon_{ia}/|y_{i,a-1}|$, and plot selected percentiles of these changes for the self-employed (Panel A) and paid-employed (Panel B).
Figure 13: Self-Employment Switching Rates by Age

A. Entry

B. Exit

Notes: The sample underlying these figures includes all individuals in the Total Sample column of Table 1. Entry rates into self-employment are shown in Panel A, for all non-self-employed and separately for the paid- and non-employed. Exit rates are shown in Panel B, with transitions to paid- and non-employed shown separately.
Figure 14: Self-Employment Switching Rates By Year

A. Entry

- To Self-employment
- From Non-employment
- From Paid-employment

B. Exit

- From Self-employment
- Paid-employed

Notes: The sample underlying these figures includes all individuals in the Total Sample column of Table 1. Entry rates into self-employment are shown in Panel A, for all non-self-employed and separately for the paid- and non-employed. Exit rates are shown in Panel B, with transitions to paid- and non-employed shown separately.
Figure 15: Differences in Past Incomes between Current and Future Switchers

A. Past Wage Income

Income Difference (thousands of 2012 dollars)

Age

B. Past Asset Income

Income Difference (thousands of 2012 dollars)

Age

See notes at the end of the figure.
Figure 15: Differences in Past Incomes between Current and Future Switchers (Cont.)

C. Past Spousal Wage Income

D. Past Adjusted Gross Income

Notes: The sample underlying these figures includes individuals with at most one observed switch between paid- and self- employment. Each panel displays the interquartiles of differences in average past incomes at each age, that is, the average wage of the switcher less the average age of peers that have similar characteristics but switch later.
Figure 16: Differences in Incomes between Switchers and Non-switchers

A. Paid- to Self-Employment

B. Self- to Paid-Employment

Notes: The sample underlying these figures includes individuals that switch at least once between paid- and self-employment. Panel A shows the income difference—averaged over three years before and after the switch—for those switching from paid- to self-employment. Panel B shows the income difference for this switching from self- to paid-employment.
Figure 17: Growth Differentials for Young Entrepreneurs

Notes: The sample underlying these figures includes all individuals born in the 1970–1975 cohorts with at least five years of self-employment experience before age 35. Growth profiles are plotted separately for those who, after age 35, remained in self-employment and those who switched to paid-employment. Growth in income for each age $a$, $Y_g(a) - Y_g(a-1)$, is computed for all subgroups $g$ and averaged using count weights, $N_g/a \sum_g N_g/a$. 

SE stayer:
$Y_{(25)} = 38,136$
$Y_{(45)} = 166,682$

SE switcher:
$Y_{(25)} = 32,685$
$Y_{(45)} = 82,680$
Figure 18: Empirical Moments, IRS versus CPS

A. Self-Employed Median Income

B. Paid-Employed Median Income

C. Self-Employed Mean Income

D. Paid-Employed Mean Income

Notes: For both the IRS and CPS samples, individuals are assigned to self-employment in a particular year/age if the absolute value of income from business exceeds $5,000 in 2012 dollars and the income from paid-employment. If these criteria are not met but income from non-business wages and salaries exceeds $5,000 (in 2012 dollars), then they are assigned to paid-employment. The sample is not balanced, and statistics are computed for each age.
Figure 19: Growth Profiles for Top and Bottom 25% Income Ranks

A. Top 25%

- Self-employed
- Paid-employed

SE incomes:
- $Y_{25} = 70,054$
- $Y_{55} = 467,788$

PE incomes:
- $Y_{25} = 48,465$
- $Y_{55} = 189,222$

Notes: The sample underlying these figures includes individuals in the attached self- and paid-employed subgroups that are ranked by income into the top 25 percent (Panel A) or bottom 25 percent (Panel B). Both panels show weighted averages of subgroup $g$ growth for each age $a$, $Y_g(a) - Y_g(a-1)$, with weights constructed from sample counts, $N_g^a / \sum_g N_g^a$. 

B. Bottom 25%

- Self-employed
- Paid-employed

SE incomes:
- $Y_{25} = 20,355$
- $Y_{55} = 16,923$

PE incomes:
- $Y_{25} = 21,728$
- $Y_{55} = 33,535$
Figure 20: Growth Differentials for Young Entrepreneurs

Notes: See Figure 17 and Section 4.3 for a description of the data. See Section 5 for a description of the structural model used to generate theoretical predictions. See Table 6 for the model parameters.
Figure 21: Model Predictions as Intangible Revenue Varied

A. Growth Differences

B. Owner Hours of Investment

Notes: See Section 5 for a description of the structural model used to generate theoretical predictions. See Table 6 for the model parameters.