LONG-RUN PROSPERITY

WORKING PAPER SERIES

HOOVER INSTITUTION

On the Nature of Entrepreneurship

Anmol Bhandari, Tobey Kass, Thomas J. May, Ellen R. McGrattan, and Evan Schulz

Working Paper 23003

June 2023

On the Nature of Entrepreneurship*

Anmol Bhandari
University of Minnesota

Tobey Kass
Office of Tax Analysis

Thomas J. May
University of Minnesota

Ellen R. McGrattan
University of Minnesota

Evan Schulz
Internal Revenue Service

September 29, 2022

ABSTRACT

This paper elucidates the nature of entrepreneurship by comparing life-cycle income profiles and outcomes of individuals who share similar characteristics but differ in their choice of self- or paid-employment. 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 entrepreneurs with at least twelve years in self-employment during our sample 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. A theory of entrepreneurial choice is developed and compared to the subsample of young entrepreneurs in our data. We find that including firm-specific investment and selection under incomplete information is necessary if the theory is to match the observed income growth profiles and switching behavior for these young entrepreneurs.

Keywords: Entrepreneurship, occupational choice, taxation
JEL classification: E13, H21

---

* 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. 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 Internal Revenue Service or the U.S. Department of the Treasury, or the National Science Foundation. All results have been reviewed to ensure that no confidential information is disclosed. All data work for this project involving confidential taxpayer information was done at IRS facilities, on IRS computers, by IRS employees, and at no time was confidential taxpayer data ever outside of the IRS computing environment.

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 life-cycle income profiles and outcomes 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 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, spousal incomes, and tax rates. We use machine learning algorithms to impute additional information such as education and skill sets. Information on the group and cohort of individuals is then used when modeling their income, which is assumed to have 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 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 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 survival bias, we separately study individuals who switch frequently between self-, paid-, and non-employment and those who are relatively "attached" in their employment status. By attached, we mean that the individual has two or fewer switches in employment status over the 16-year sample period with no years of non-employment. Our measure of income for the self-employed is the sum of incomes from proprietorship (reported on Form 1040, Schedule C), partnerships (reported on Form 1065, Schedule K-1), S corporations (reported on Form 1120-S, Schedule K-1), and own-business compensation (reported on Form W-2). Income for the paid-employed is wage income (reported on Form W-2) less any own-business compensation.

Comparing income profiles for the full sample of attached employees and entrepreneurs—before making any adjustments for fringe benefits or taxes—we find that growth in employee wage earnings decline across the life cycle, while growth in entrepreneurial income remains persistently high

[1] See Parker (2018) for a comprehensive review of the literature.

until mid-career and then gradually declines. The mean incomes start around the same level at age 25: \$34.6 thousand (in 2012 U.S. dollars) for the paid-employed and \$42.5 thousand for the self-employed. However, by age 55, the self-employed are earning more than twice the paid-employed, roughly \$202.1 thousand versus \$96 thousand. If we decompose the differences in income growth around age 33—when they peak—between the self- and paid-employed 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 profiles are highest during the mid-30s, suggesting that business owners make initial investments to build a business (as in McGrattan and Bhandari, 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 more hump-shaped growth profile that we observe.

To gain a better understanding of entrepreneurial choice, we analyze entry into and out of self-employment over the life cycle and over our sample period, constructing statistics that are informative for testing alternative theories. For our sample, exit rates decline significantly over the life cycle, with experimentation in entrepreneurship occurring in the early years, 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).) We also find that the new entrants have higher earnings after the switch when compared to their observationally similar counterparts, which suggests that entrepreneurs are driven by pecuniary motives.

After documenting 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, a customer base—and growing to an optimal size. There are risks in self-employment and young entrepreneurs start with little to no financial assets 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. 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, then occupational choices are made quickly. In this 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 be scaled to its optimal size immediately. In this case, the model cannot rationalize persistent differences in income growth when comparing profiles for entrepreneurs that continue in business and those that exit.

An important by-product of our work is a longitudinal database 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, while we can study the typical dollar earned in self-employment. What our analysis shows is that this typical dollar is earned by those with incomes in the top 25 percent that are relatively attached to self-employment.

2 Data

In this section, we describe our main sample drawn from U.S. administrative tax records.[2] We start with details of the data source and definitions of income for the paid- and self-employed. We then describe algorithms that infer education and skill levels using information on educational tax credits and occupation.

2.1 Sample

The largest potential sample contains records in the IRS Statistics of Income (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.[3] For each individual there are rows, one for each year, and columns recording demographic information from the Social Security administration (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.

[2]Replication codes and detailed documentation are available at the IRS.

[3]See Chetty et al. (2018) for full details on this database. We remove any person from our sample who died prior to 2015.

The information on wages and salaries reported on individual W-2s for employees and Schedule-C incomes for self-employed proprietors are readily available in the Databank. For owners' pass-through businesses—partnerships and S corporations—we merge in information from Schedule K-1 filings attached to Form 1065 and 1120-S, respectively. The Schedule K-1 data is available since 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 individuals that used Form 1040A or 1040EZ in any sample year from our main sample and investigate them separately in our sensitivity analysis. Roughly 40 percent of all individual tax filers use these simpler forms each year. Our sample also excludes owners of Subchapter C corporations that are taxed as their own legal entities.

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 127 million individuals for 16 years (that is, 2 billion person-year observations).[4] 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 restrictions narrows our sample to roughly 65 million individuals. Details of the imputations are provided later in section 2.3 below.

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 1120S, 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).[5] The second is a measure of *paid-employment income*, which is the wages and salaries paid by businesses they do not own (Form W-2, Box 1). 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.

[4]The full SOI Databank sample over 2000–2015 has 6.7 billion observations and 3.1 billion for ages 25-65 if we include individuals that are not in our sample all years.

[5]In our baseline, we omit capital gains as source of self-employment income. Indeed some part of the capital gain could arise from sale of intangible assets that reflect entrepreneurial investments and should be considered on par with self-employment income. Given our main findings, we view omitting capital gains as a conservative stance. Decomposition of capital gains into labor- capital-related components is ongoing work with the IRS, but outside the scope of the current paper.

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.

Definition 1. An individual-year is classified as self-employed (SE) if any of the following is true: (i) the individual's share in the business times number of employees of the business is larger than 1; (ii) the absolute value of their self-employment income is greater than the absolute value of their paid-employment; or (iii) the individual's share of gross profits, that is, receipts minus cost of goods sold, are in excess of the individual's paid-employment income.

The first criteria is added because hiring employees is indicative of owner attachment to self-employment. The second criteria uses an absolute value on self-employment income because young entrepreneurs incur significant expenses when building up their businesses and the businesses earn losses. The third criteria allows for the fact that many successful business owners pay themselves little 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), Garin, Jackson, and Koustas (2022), and DeBacker, Panousi, and Ramnath (2022), who all use IRS data to study business incomes. Smith et al. (2019) classify all individual recipients of K-1 as self-employed. Under our definition about 37 million individual-year K-1 recipients would not be classified as self-employed. These are cases in which an individual 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. Garin, Jackson, and Koustas (2022) focus on Schedule SE filers. This is not suitable for our analysis because it misses entrepreneurs who make losses and S corporation owners that do not file Schedule SE. This is a significant fraction of business owners. DeBacker, Panousi, and Ramnath (2022) use a panel that tracks tax filers for up to 32 years using the SOI sample from 1987 that is the basis of the published statistics. While this has the benefit of being a long panel, the number of self-employed individuals that are studied shrinks down to about 2000 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.

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 W-2 earnings 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 PE or SE but missing in the SOI databank, we use auxiliary data from the SOI databank in the year before and after the missing observation, as well as incomes reported on the individual's Form W-2, Schedule C, or Schedule K-1, and Form 1040 if any of these filings are available. Consider wage earners first. If amounts on Form W-2 and Form 1040 are available and the same—say, because the individual is single or a spouse does not earn wage income—we can use that information to fill in our missing data. If they differ and we can confidently attribute the income to the individual alone, we use the Form 1040 information if greater by $1,000. If we have missing observations on all forms for the year, but the SOI databank reports the same level of income in the prior and subsequent year, we use that level of income to fill in for our missing observation. In cases with married couples filing jointly, we compare the combined W-2 incomes to the amount on Form 1040. If they match, we have the information we need to fill in the missing observation. If not or if the W-2 information is unavailable, we use available data from the prior or subsequent year to divy the Form 1040 income into pro-rata shares for the two spouses. In the case of business owners, we follow the same procedures with the Schedule C and Schedule K-1, which are forms specific to individuals, as we did with Form W-2 for individual wage earners.

In the first column of Table 1, we provide summary statistics for our main sample. There are 65 million in the full sample. They earn an average of $53.5 thousand (in 2012 dollars) in combined paid- plus self-employment income, with a range across the distribution from $6.7 thousand at the 10^{th} percentile to $99.7 thousand at the 90^{th}—roughly a factor of 15. The mean income from paid-employment—if we average across all individual-year observations in the sample that are characterized as PE under Definition 2—is $45.8 thousand, which is more than 6 times the mean of self-employed income for our sample. 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 of SE income, indicating a substantial right-skewness of the self-employment income distribution for the whole population.

2.3 Imputations for Skill and Education

A large empirical labor literature focuses on skills and education as determinants of income. Such information is not directly recorded on the main tax forms and therefore not present in the SOI Databank. We use machine learning algorithms that are trained on auxiliary survey data to impute measures of skills and educations. We later use these measure to form groups and to do our analysis within groups.

Skills After signing and dating the tax form, individual tax filers and their spouses are asked to enter their occupation. The occupation information is only available for e-filed returns starting in 2013. For the sample of individuals born between 1950 and 1975, there are 90 million individuals that e-filed at least once in the years 2013–2015. We are able to assign skill values to the subset of 65 million individuals in our main sample in two steps.

First, there are 56 million that provide occupations responses that can be mapped directly to a standard occupational classification (SOC) code.[6] 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 detailed 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 and construct indices—keeping the top three (orthonormal) components and ensuring that occupations requiring mathematics reflect "cognitive" skills, occupations requiring social perceptiveness reflect "interpersonal" skills, and occupations requiring mechanical knowledge reflect "manual" skills. After applying this method, we map SOC codes in the IRS dataset to skill values. We use Z_{ijt} to denote the skill value of individual i along dimension j (say, cognitive, interpersonal, or manual) in year t.

Second, there are 9 million individuals with responses that cannot be directly mapped to an SOC code.[7] For these cases, we developed a classifier using the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) for the years 1995–2020. In this case, we run the regression

$$Z_{ijt} = \beta_{jt} X_{it} + \epsilon_{ijt}, \tag{1}$$

where ϵ_{ijt} is a disturbance term, and the variables included in X_{it} are as follows: gender; annual pre-tax wages and salaries; positive business income (which is equal to 0 if income is negative); negative business income (which is equal to 0 if income is positive); marital status; number of children (with separate variables for none, 1 child, and so on, up to 9 or more); 5-year birth cohort; and 2-digit NAICS industry code. Skill value estimates are the fitted values from the previous regression.[8] We finally reduce Z_{ijt} along the time dimension by taking the highest skill value that we have imputed.

[6] We thank Raj Chetty and his team for providing us with a mapping between the strings and the SOC codes.

[7] For instance, a lot of business owners fill in "self-employed," which is not a valid SOC code.

[8] We also experimented with probit and logit specifications but the linear classifier outperformed other specifications in our validation metrics.

Education Imputation As with skills, we develop an education classifier that predicts college enrollment. The source data used to train the classifier is similar to the previous classifier for occupation, that is, the ASEC and CPS for the years 1995–2020. We define an individual as being "educated" if they have at least an associate's degree (and thus include bachelor's, master's, professional school, and doctorate degrees). All others are considered "not educated." For each year t, we run the following regression:

$$\Pr(E_{it} = 1|X_{it}) = \text{CDF}(\beta_t X_{it}), \tag{2}$$

where $E_{it} = 1$ if the individual is educated and 0 otherwise. The function CDF in (2) 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, 1 child, and so on, up to 9 or more); 5-year birth cohort; SOC minor occupation code; and 2-digit NAICS industry code.[9] When we used 90 percent of our CPS sample each year to train the classifier and 10 percent to validate the predictions, we were able to correctly predict the education level with 75 to 80 percent accuracy.[10] Coefficients from the CPS-trained classifiers then are used with microdata from the IRS to impute an education indicator for all tax filers in our sample of 25 to 65 year olds that do not have a 1098-T or an occupation string indicating that they are a student.[11]

In Table 1, we report that the fraction of individuals categorized as educated for the main sample in the first column under "skills." The imputation results indicate that 52.1 percent of the main sample is classified as educated, which is slightly higher than the 47.2 percent estimate for cognitively skilled. More individuals are categorized as interpersonally skilled (56.1 percent) than manually skilled (32.3 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

[9] Some IRS tax filers do not have a valid NAICS code and do not have a 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.

[10] As in the case with skills, we experimented with linear and logit specifications but the probit classifier outperformed other specifications in our validation metrics. The details of the classifier and other related diagnostic tests are included with documentation for the codes archived at the IRS.

[11] All variables in X_{it} are available in the IRS data, although the IRS occupation field is only available for tax years 2013 and later and is coded as a string rather than as a SOC.

the challenges to accurately measure the age profile of income and then describe how we use the novel features of our data with an econometric method to overcome these challenges.

3.1 Some Challenges

A natural starting point for measuring returns to employment is to specify a Mincer-type 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).) There are several concerns with this approach. First, differences in average incomes could be driven by selection—we are simply comparing individuals that differ in their latent characteristics, preferences, or shocks. For instance, suppose that after controlling for observables, income in paid-employment is less volatile than self-employment. These outcomes would reflect latent differences in risk preferences but would not reflect differences in investment opportunities. Second, differences in average incomes could be driven by survival. For instance, suppose adverse shocks cause individuals to switch from self-employment to paid-employment. In this case, an increasing pattern in the average lifecycle income for self-employed individuals would reflect favorable changes in the composition of those who remain self-employed.

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 they are latent or observed—as well as calendar time 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 selection. We then separately study income-age profiles of individuals who are relatively attached by narrow gender/skills/industry categories and income-age profiles of individuals who are less attached, transiting 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 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 survival bias.

We start with some notation. Let $i \in I$ be a set of individuals; $t \in \mathcal{T} = \{t_0, t_0 + 1, \ldots t_0 + T\}$ be a set of calendar dates; $c \in \mathcal{C} = \{c_0, c_0 + 1, \ldots c_0 + C\}$ be a set of birth years; $a \in \mathcal{A} = \{a_0, a_0 + 1, \ldots a_0 + A\}$ be a set of ages; and $g \in \mathcal{G}$ 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 $\beta : \mathcal{G} \times \mathcal{T} \to \mathcal{R}$ and $\gamma : \mathcal{A} \times \mathcal{G} \times \mathcal{C} \to \mathcal{R}$ that capture time and age 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 $\gamma^{a(i,t)}_{c(i),g(i)}$ to be the values associated with an individual-time pair (i, t). Consider the following specification for income

$$y_{i,t} = \alpha_i + \beta_{g(i),t} + \sum_{a=a_0}^{a=a(i,t)} \gamma^a_{c(i),g(i)} + \epsilon_{i,t}. \tag{3}$$

where $\epsilon_{i,t}$ is a disturbance term for individual i at date t.

The model for income in equation (3) 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 such as risk-aversion 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 $\left\{\gamma^a_{c,g}\right\}$ 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 skill, industry, and demographics.

It is well-known and easy to see that one cannot separately identify β and γ 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 assumption.

Condition 1. Age-effects are the same across cohorts, that is, $\gamma^a_{c,g} = \overline{\gamma}^a_g$.

It is worth pointing out that while we impose age effects across cohorts to be 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, α_i. Condition 1 allows us to exploit the overlapping structure of our data to separate out age effects from time effects.[12] In our main sample, we see 26 cohorts (birth years 1950–1975), across 41 ages (25–65) across 16 calendar years (2000–2015). Thus, there is a significant overlap of cohorts over time.

Next, we derive the formulas needed to implement the estimation procedure. Let Δ be the time difference operator so that $\Delta x_t = x_t - x_{t-1}$. Apply Δ to equation (3) to obtain

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

To estimate the age and time effects, we propose the following least squares problem:

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

[12] As a robustness check, we relaxed Condition 1 to allow age effects to vary across 5-year and 10-year bins of cohorts.

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

$$\begin{aligned}
\overline{\Delta y}_{g,t} &= \frac{\sum_{i\in\mathcal{I}:g(i)=g}\Delta y_{i,t}}{\sum_{a\in\mathcal{A}} N^a_{g,t}}\\
\overline{\Delta y}^a_g &= \frac{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}:a(i,t)=a,g(i)=g}\Delta y_{i,t}}{\sum_{t\in\mathcal{T}} N^a_{g,t}}
\end{aligned}$$

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

$$\overline{\gamma}^a_g = \overline{\Delta y}^a_g - \sum_{t\in\mathcal{T}}\left(\frac{N^a_{g,t}}{\sum_{j\in\mathcal{T}} N^a_{g,j}}\right)\underbrace{\left\{\overline{\Delta y}_{g,t} - \sum_{k\in\mathcal{A}}\left(\frac{N^k_{g,t}}{\sum_{\ell\in\mathcal{A}} N^\ell_{g,t}}\right)\overline{\gamma}^k_g\right\}}_{\Delta\beta_{g,t}}. \tag{4}$$

Equation (4) expresses $\left\{\overline{\gamma}^a_g\right\}$ as linear combinations of two summary statistics of data, $\left\{\overline{\Delta y}^a_g\right\}$and $\left\{\overline{\Delta y}_{g,t}\right\}$ with weights $\left\{N^a_{g,t}\right\}$. Specifically, the age effects for some age a are given by the average income growth $\overline{\Delta y}^a_g$ 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,

$$\frac{N^a_{g,t}}{\sum_{a\in\mathcal{A}} N^a_{g,t}} = \frac{\overline{N}^a_g}{\sum_{a\in\mathcal{A}}\overline{N}^a_g}, \tag{5}$$

where $\overline{N}^a_g = \sum_{t\in\mathcal{T}} N^a_{g,t}$. With some algebra, we can show that $\overline{\gamma}^a_g = \overline{\Delta y}^a_g - \overline{\Delta\beta}_g$, where $\overline{\Delta\beta}_g = \sum_{t\in\mathcal{T}}\Delta\beta_{g,t}/T$ is the average of time effects across time 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.[13]

We make two more observations about equation (4). First, the age effect $\overline{\gamma}^a_g$ 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 $\left\{\overline{\gamma}^a_g\right\}$ uniquely. We impose the following condition.

[13]In our sample, we have a balanced panel and, therefore, the mean age is necessarily increasing in calendar time as the population is aging.

Condition 2. The average time effect satisfies

$$\frac{\overline{\Delta\beta}_g}{\overline{y}_{g,t_0}} = \frac{\mu_g \sum_t (1+\mu_g)^t}{T} \tag{6}$$

for some pre-determined constant μ_g, where $\overline{y}_{g,t_0} = \sum_{i\in\mathcal{I}:g(i)=g} y_{i,t_0} / \sum_{a\in\mathcal{A}} N^a_{g,t_0}$ 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. In particular, we do not need to take a stand on the differential effects of aggregate shocks on groups. This is especially helpful in our sample given the severe economic downturn in 2008 and 2009.

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 several time-invariant characteristics that we call *subgroups.* In our case, there are 35,117 subgroups with value assignments. In this section, we provide a summary of the subgrouping.

Given our interest in the returns to entrepreneurship, the most relevant characteristic is how attached an individual is to self-employment as opposed to paid- or non-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 survival bias, we analyze income profiles by separately studying 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 one of the three employment categories. An individual is labeled *attached* if we observe two or fewer changes in employment status during the sample.

In Table 1, we report counts and characteristics of the relatively attached individuals in the second and third columns. There are 35.4 million in paid-employment, which is about 54.4 percent of the individuals. They earn 66.8 percent of total income, 77.2 percent of all paid income, and roughly 4.1 percent percent of entrepreneurial income. The attached self-employed, at 1.9 million individuals, comprise only 2.9 percent of the individuals in our main sample, but 8.4 percent of total income and 52.5 percent of entrepreneurial income.

Separate results are reported for groups that do more switching in employment status. The *almost attached* groups have the same employment status for twelve 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. The *mostly switchers* have twelve or more years in either self- or paid-employment—without an intermediate spell of non-employment—and experience at least five or more years in both types of employment. This group is similar in size to the attached self-employed and will be used to gain insight into motivation for entering and exiting self-employment. The last category is *any non-employment* which 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 roughly one-third of the sample but is different in many ways from all other subpopulations: on average, these individuals earn significantly less, are primarily low-skilled, and have a higher concentration of singles and women than any other category.

In addition to employment attachment, we use other observables to group individuals. The subgroup *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 indicator is above the 0.5 cutoff and 0 if not. *Industry* has 20 values for the 2-digit NAICS codes. 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 3 values: "1950" if born between 1950 and 1959; "1960" if born between 1960 and 1969; and "1970" if born between 1970 and 1975. We assume that individuals are included every year of our sample. Thus, as an example, the "1950" cohort would contain individuals with ages between 41 and 50 in the first year of the sample. These same individuals would be between 56 and 65 in the last year of the sample.

The industry and demographic summary statistics for our main sample are reported in the bottom rows of the first column of Table 1. We have information on 2-digit industries of the employer in the case of employees and the primary business in the case of entrepreneurs. The demographic data show that 50.7 percent of the sample is male. Most are married for a majority of years they are in the sample—about 67.6 percent—and most have children—about 82.5 percent. The median birth year for our sample is 1963. The table also shows distributions of income across the different subgroups and highlights the fact that self-employed incomes are significantly higher than the paid-employed counterparts across most of the distribution. This is true despite little difference in shares by education or skill. The main differences appear to be in the sectors where employment occurs and in the demographic variables. Self-employment is concentrated in construction, professional services, and health care and the entrepreneurs are mostly male and married.

4 Results

In this section, we report on the estimated income profiles for our sample and investigate determinants of self-employment. We emphasize those empirical results most relevant for distinguishing between theories of occupational choice. We start by summarizing cyclical growth patterns over our sample using estimated time effects, $\{\beta_{g,t}\}$, with a particular focus on the impacts during the 2008–2009 downturn. We then study income and growth profiles over the life cycle using estimated age effects, $\{\overline{\gamma}_g^a\}$, for major subpopulations. We compare profiles for the self-employed with those for the paid-employed and find significant differences. We document that certain subpopulations drive these differences as they account for most of the self-employed incomes. We then study entry into and exit out of self-employment to better understand the determinants of entrepreneurship. Finally, we compare findings with the empirical literature with the aim of informing theory development.

4.1 Entrepreneurial Incomes

We start with our estimates of time and age effects that are used to construct income profiles for the self- and paid-employed.

4.1.1 Cyclical Growth

In Figure 1, we plot the time effects, $\{\beta_{g,t}\}$, for individuals that are relatively attached, either to self- or to paid-employment, and thus to their employment status for at least twelve years.[14] The values are reported in thousands of 2012 U.S. dollars and displayed for tax years 2001 to 2015. As expected, there is a decline in growth during the 2008–2009 recession, with paid-employed growth in incomes falling to about $-\$1.1$ thousand dollars and self-employed growth in incomes falling to about $-\$15.5$ thousand. Interestingly, both see improvements in growth by 2010. The estimates show much larger variation in self-employed growth, although the magnitudes of changes in incomes should be compared to estimated age effects, which we discuss next.

4.1.2 Life-cycle Growth

In Figure 2, we display the age effects for the same sample of attached self- and paid-employed used in Figure 1. Panel A shows the average integrated incomes, that is, for each age $a \geq 25$, average income for 25 year olds in group g plus $\sum_{j=26}^{a} \gamma_g^j$. The data are deflated with the BEA's personal consumption expenditure price index and reported in thousands of 2012 dollars. They are reported pre-tax and exclude fringe benefits. The dots are the point estimates and the bold lines are a third-order polynomial fit. Panel B shows the age effects $\left\{\gamma_g^a\right\}$. We also report the incomes at age 25 and 55 from Panel A as we do in growth profile figures for subpopulations shown later.

[14] In all results reported, we exclude the top and bottom 0.01 percent outliers in adherence with IRS disclosure rules when considering subgroups of the population. Our main aggregated results are not affected if we include them.

Heterogeneity in age-effects for the attached As the figure shows, the age effects differ substantially across the groups. Despite earning roughly the same early on, the self-employed incomes are significantly higher by age 55 than their paid-employed counterparts. Thus, while a $15.5 thousand loss in the 2008–2009 recession is significant, we need to compare it to the average annual income of $154.4 thousand for the self-employed rather than an average annual income of $65.6 thousand for the paid-employed. Another striking difference is the life-cycle growth patterns shown in Panel B. Growth rates in paid-employment decline across the life cycle, while growth rates in self-employment do not. The attached self-employed have persistently high average growth in incomes—in the range of $4 thousand to $8.7 thousand annually—for ages between 25 and 40. Even after that, the self-employed average growth rate remains significantly higher than that for paid-employees and, by age 55, the self-employed have an average income over $202.6 thousand—more than twice that of the paid-employed group and too large to be due to differences in taxes or benefits.

Next, we repeat the exercise for subpopulations 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 growth profiles for women in Panel B. Recall that men account for most of the attached self-employment sample, roughly 53.4 percent, but only 82.4 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 much higher than for their paid-employed counterparts, although less hump-shaped than that for men. 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. Peak income for self-employed women—at roughly $143.2 thousand—is about 88.5 percent of the male average. Peak income for paid-employed women is close to $76.9 thousand—about 84.3 percent of the male average.

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 79.1 percent of the relatively attached self-employed population and 70.3 percent of the relatively attached paid-employed. The mostly unmarried have similar patterns in growth, but the incomes of the self-employed are $96.2 thousand higher by age 55.

Education is another important characteristic when considering returns to experience. After applying our classifier to categorize individuals in the "likely" or "not likely" to have attained a college degree, we recompute slopes and construct average growth profiles. The results are shown in Figure 5. The differences are striking. Average annual growth rates of the likely educated self-employed reach a peak of \$13.9 thousand at age 34—roughly \$10.3 thousand more than the educated paid-employed—while the growth rates of the non-educated groups are never higher than \$2.1 thousand, regardless of employment status. For the not likely educated, there is little difference in magnitudes or growth profiles over the life cycle across employment categories.

With the O*Net data and our occupation strings, we can study returns to self- or paid-employment for people with different skillsets. 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 between likely- and not-likely-educated are much starker than differences between cognitive and not-cognitive—regardless of occupation choice. In the case of the self-employed, the cognitive have average incomes that peak around \$211.3 thousand while the non-cognitive have average incomes that peak close to \$189.1 thousand.

In Figure 7, we show growth profiles by interpersonal skill. While not as stark as our results by education, there are still large differences for the skilled and unskilled in this case. For example, 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 or the paid-employed. However, the average income of the interpersonally unskilled at age 55 is not as low as that of their uneducated counterparts.

In Figure 8, we show growth profiles by manual skill. In this case, the self-employed have higher growth regardless of skill, but the higher earners are in occupations that do not require manual skills. Even so, by age 55, those self-employed with manual skills earn around \$100.7 thousand on average, which is comparable to the mean income of the non-manually skilled in paid-employment.

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 relatively-attached employment status that are in professional services (NAICs 54) and construction (NAICs 62). Together these sectors account for one-third of the relatively-attached self-employed population. Here, we see the same patterns as before: large differences between the income growth rates of the self- and paid-employed and hump-shaped growth profiles for entrepreneurs. The main difference between these results and the averages in Figure 2B is in the levels: by age 55, the average self-employed income in professional services is \$296.9 thousand whereas the average in health care is \$271.9 thousand.

The results shown thus far are for relatively broad categories and include people with a wide range of characteristics. Because there are thousands of possible cuts of the data, 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, the largest gap in income growth occurs between ages 33 and 34, at roughly \$6 thousand. 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 the attributes of the group. For example, the group contributing the most to the growth differential is married males in the health care sector (NAICS 62) that are educated and have both cognitive and interpersonal skills. This group of primarily medical doctors contributes 15.4 percent to the growth gap, with the self-employed among them experiencing average annual growth in income around \$20 thousand during their 30s—roughly 1.6 times the growth of their paid-employed colleagues. As Table 2 makes clear, there are a small number of sectors that matter for our results: health care, professional services, construction, finance, and wholesale trade.

Variability and persistence of $\{\Delta\epsilon_{i,t}\}_{i,t}$ Besides the time- and age-effects, our estimation procedure yields $\{\Delta\epsilon_{i,t}\}_{i,t}$ whose statistical properties are of interest. Here we focus on the variability and persistence of $\{\Delta\epsilon_{i,t}\}_{i,t}$ that are relevant for discussions of entrepreneurial risk-taking and more generally for the larger debate on earnings inequality in the U.S.

To provide a sense of the potentially risky nature of entrepreneurship, we plot percentiles of growth rates in Figure 10 for primary working ages 26 to 55. Here, we compute the change in income between ages a and $a-1$, divided by the absolute value of income at age $a-1$ for all individuals in our samples of attached self- and paid-employed. Panel A has results for the attached self-employed and panel B has results for the attached paid-employed. For both groups, the percent changes of income is most disperse 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. This feature is relevant for theories that assume income volatility grows over the life cycle for entrepreneurs when compared to similar paid-employed peers, and this income risk must be compensated with higher returns in later years.

To understand the persistence of earnings, we compute income transition matrices separately for individuals in our samples of attached self- and paid-employed. These are displayed in Table 3 along with the distributions of income. Each element of the matrix is the share of individuals that starts the year in a particular income bin listed at the top of a column and then experiences a shock—one that we pick up residually as $\Delta\epsilon_{i,t}$ in our estimation procedure—such that the resulting income bin before accounting for time and age effects is listed at the left of the row. Comparing the two matrices, we find significantly less persistence for the self-employed when compared to the paid-employed. This finding is consistent with increased mass in the upper and lower income ranges for the self-employed.

Mostly-switchers and any-non-employed groups Thus far, we have investigated results for the attached subpopulations but a significant fraction of self-employed income is earned by our mostly-switchers and any-non-employed subgroups. (See Table 1.) In Figure 11, 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. Panel B shows the growth profile comparison between the attached self-employed and the non-employed. The non-employed profile has the same pattern as that of the paid-employed, although growth is lower at all ages.

To summarize the findings based on life-cycle profiles, we find that entrepreneurs with at least twelve years in self-employment during our sample have significantly higher average incomes and steeper, more persistent, income growth than the attached paid-employed, the mostly-switching, and the any-non-employed. The fact that growth peaks later for the successful entrepreneurs is suggestive that they are making firm-specific investments early on, investments that pay off later in their career in terms of higher growth rates and higher incomes. This growth pattern will inform the theory we develop later.

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 income, and a dominant role for pecuniary motives to start self-employment.

We focus on switching rates between the three 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 12, 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 1.3 percent to 2.5 percent and is mildly increasing in age. As is clear from the figure, most of the rise is due to entry from non-employment. Exit rates out of self-employment are shown in Panel

B. The overall rate is high and strongly declining, starting at 37.5 percent and dropping to 17.2 percent by the end.[15] This declining hazard rate is suggestive that experimentation and learning about the potential gains to entrepreneurship occurs 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 13, we plot the entry and exit rates by tax year. In this case, we purge an age effect that arises due to the aging population over our sample period.[16] 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, 2006.)

To better understand the motives and impediments to switching, we compare past labor income and asset income for switchers to that of comparable non-switchers. Our first exercise compares past income—averaged over three years prior to the switch—for one-time switchers into self-employment to past income of those who do not switch, but share the birth year, gender, industry, and lagged employment status. In Figure 14, we plot the interquartiles of this difference in past income by age. A positive value indicates switchers have higher past income than the non-switchers. We see that early switchers have similar past incomes to non-switchers and over time the gap becomes larger and more favorable for the switchers. The median difference is roughly zero at age 28 and goes up to zero by age 55. If we focus on those who switch from paid- to self-employment, we find similar patterns but larger differences. From this exercise, we conclude that switchers are positively selected on past productivity.

Next, we compare asset incomes of switchers and non-switchers. Since labor and asset incomes are generally correlated, we isolate the role of assets for entrepreneurial choice by comparing switchers to non-switchers who not only share birth year, gender, industry, lagged employment status, but also the percentile of past income. In Figure 14 Panel B, we plot the distribution over age of past asset income of the switchers less an average of past asset income of the non-switchers they are paired with. Call this difference the excess asset income. For most switchers, we find the differences to be negative and small when compared to wage incomes. The median excess asset income ranges from roughly zero at age 28 and to close to $-\$4.2$ thousand at age 55. These findings hold up even if we focus exclusively on those in paid-employment prior to the switch. From this exercise, we conclude that switchers are negatively selected on liquidity.

[15] Although not shown here, there is a also distinct gender gap in the overall entry and exit rates. Entry rates for women are in the range of 0.8 percent to 2.2 percent and, relative to men, their exit rates are roughly 6.3 percent lower at all ages.

[16] 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.

To distinguish between pecuniary and nonpecuniary motives for switching, we ask if growth in income improves or worsens following a switch. Modest or negative growth in incomes after a switch would be consistent with a large role for non-pecuniary motives driving the switch. To minimize the concern of selection in the set of switchers, we compare switchers at a particular age to a set of switchers who eventually switch but did not switch at that age. In Figure 15, we plot the average annual growth differential following a switch—along with the interquartile ranges—weighted by subgroup counts, for ages 32 to 50. In the case of those switching from paid- to self-employment, there are declines for a little over half of the switchers when compared to peers, but the average annual growth differential is high, reaching peaks roughly \$18.5 thousand. This suggests that higher earnings in self-employment is a dominant motive for many that are switching. For those switching from self- to paid-employment, the mean and median are barely different and on the order of \$1.9 thousand for most ages. But here again, there are a large number of individuals with higher post-switch income when compared to their peers.

Finally, because we find steep hazard rates at early ages, we further investigate switching behavior of individuals in the youngest cohort of our sample. More specifically, we next ask if there are differences between income growth profiles of 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 (born 1970–1975) with at least five years of self-employment experience prior to age 35. In Figure 16, 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 striking 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. 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.

4.3 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. There is a large literature that 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) and Kartashova (2014) with the Survey of Consumer Finances.[17] This literature has been extremely influential in promulgating our understanding of entrepreneurship

[17]For a comprehensive set of references, see Parker (2018).

and motivating theories that can be used for policy analysis. In this section, we relate our findings to these studies—delineating points of agreement and points of disagreement.

Since survey data has issues related to top-coding and small samples, most research on entrepreneurship has focused on the median incomes of the self-employed. Thus, to relate our findings to those based on surveys, we start by plotting median and average self-employment and paid-employment incomes by age for all individuals in our sample. In Panel A of Figure 17, we see a familiar result: median self-employment income is below the paid-employment income—by roughly $15.9 thousand—at 35. This is consistent with an abundance of survey evidence that finds a self-employment "discount."[18] This finding has solidified a view that self-employed individuals must be earning large nonpecuniary 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) and Hall and Woodward (2010) that emphasize low returns relative to the risk in self-employment.

In Panel B of Figure 17, we see that the average income for the self-employed in our sample is $45.5 thousand above the average paid-employed at the peak. We should note that there are several advantages of the econometric approach laid out in Section 3.2 over working with these simple averages. Our analysis exploits the unique long panel that administrative data provides to separately analyze groups that differ with respect to their attachment to self-employment and is therefore less prone to selection and survival bias. As a result, we have a better perspective into the nature of self-employment and features that theories of entrepreneurship need to have. However, neither the simple averages nor the earlier average age effects in Figures 2 and 11 can be compared to survey data. Survey top-coding rules out comparisons to simple averages and limited longitudinal information along with small samples rules out comparisons to our estimated age effects.

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 explore this further, we introduce a group based on ranking individuals on 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 ignored their employment status (PE/SE) before assigning them a rank. After ranking them, we bin individuals into five quintiles. Table 4 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 80 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 18, we plot

[18] Although Levine and Rubinstein (2016) 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 incomes of the self- and paid-employed at higher quantiles of the distribution where the incorporated owners would naturally be.

the growth profile for those in the top 25 ranks by the four groups that we defined. Compared to paid-employed, we see high growth rates not only in the attached self-employed category but also in the mostly switchers and any-non-employment categories. Our take away from this is that most of 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 paid-employed. In other words, we find patterns that are quite different from those emphasized by the current literature.

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 census data such as Decker et al. (2014). 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, while 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 account for two-thirds of self-employment income. The entry rates into the self-employed-with-employees group decline from roughly 0.85 percent before 2008 to levels around 0.65 percent thereafter—a decline of about 24 percent. This is in line with the findings of Decker et al. (2014) and Bayard et al. (2018) that show a comparable decline in business start-ups.

Where we differ with the literature is in our conclusions concerning selection into entrepreneurship. A common finding in the literature is that individuals entering self-employment have lower past labor incomes when compared to peers that are similar but did not enter. As Evans 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.[19] 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 14.) Since we also find that there are strong pecuniary motives for switching and large ex-post returns to choosing self-employment, there may be top-coding issues with the survey data that are 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

[19] See also Alba-Ramirez (1994), Rissman (2003), and Rissman (2007).

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 and Fairlie.

5 Implications for Theory

In this section, we postulate an optimization problem for a theoretical entrepreneur making firm-specific investments and compare our predictions for incomes and growth to empirical counterparts discussed above. Following Bhandari and McGrattan (2021), we model the firm-specific investments as 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 also assume that the 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. We use this theoretical laboratory to generate predictions about income and growth profiles of young entrepreneurs—those that are relatively attached and those that ultimately switch.

5.1 Model

Each period entrepreneurs have to 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 κ, ability in paid-employment ϵ, latent ability in self-employment z, and years of experience j. Because z is latent, the predicted mean μ 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_κ, and expenses, e, which are inputs in the technology $f_\kappa(h_\kappa, e)$. Production requires the stock of intangible assets, κ, 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:

$$\begin{aligned} V_k(s) &= \max_{c,h_y,h_\kappa,k,n,e} \{U(c,\ell) + \beta EV(s')\} \qquad (7) \\ \text{subject to } a' &= (1+r)a + pe^z f_y(\kappa, h_y, k, n) - (r+\delta_k)k - wn - e - c \\ \kappa' &= (1-\delta_\kappa)\kappa + f_\kappa(h_\kappa, e) \\ \ell &= 1 - h_y - h_\kappa \\ a' &\geq 0 \end{aligned}$$

and additionally processes for updating ϵ, z, and μ 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 δ_k for tangible capital and δ_κ 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 η_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 η_j is normally distributed with mean 0 and variance σ_η^2 that is constant for all j. Let μ_j and σ_j^2 denote the predicted mean and variance of the latent productivity conditioned on past signals, that is:

$$\begin{aligned} \mu_j &= E[\bar{z}_0 | z_0, z_1, \ldots, z_{j-1}] \\ \sigma_j^2 &= \text{var}[\bar{z}_0 | z_0, z_1, \ldots, z_{j-1}], \end{aligned}$$

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 μ_j and variance $\sigma_j^2 + \sigma_\eta^2$. In this case, the means and variances evolve with one more year of experience as follows:

$$\begin{aligned} \mu_{j+1} &= \mu_j + \frac{\sigma_j^2}{\sigma_j^2 + \sigma_\eta^2}(z_j - \mu_j) \\ \sigma_{j+1}^2 &= \frac{\sigma_j^2 \sigma_\eta^2}{\sigma_j^2 + \sigma_\eta^2}. \end{aligned}$$

The next period value in (7) 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 ϵ which we model as a Markov chain with transition probability $\pi(\epsilon'|\epsilon)$. Then the value of sale is given by

$$\begin{aligned} V_s(s) &= \max_{c,h_y,k,n} \{U(c,\ell) + \beta \sum_{\epsilon'|\epsilon} \pi(\epsilon'|\epsilon) V_w(s')\} \\ \text{subject to } a' &= (1+r)a + pe^z f_y(\kappa, h_y, k, n) - (r+\delta_k)k - wn + p_\kappa \kappa - c \\ \ell &= 1 - h_y \end{aligned} \tag{8}$$

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

$$\begin{aligned} V_w(s) &= \max_{c,h} \{U(c,\ell) + \beta \sum_{\epsilon'|\epsilon} \pi(\epsilon'|\epsilon) V_w(s')\} \\ \text{subject to } a' &= (1+r)a + w\epsilon h_y - c \\ \ell &= 1 - h_y \end{aligned} \tag{9}$$

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); their estimates were chosen to align theoretical predictions of their model with aggregate data from the Bureau of Economic Analysis's (BEA) national accounts and micro data from the Census's Survey of Business Owners (SBO) and the Pratt's Stats database of brokered business sales. Additionally, to ensure that some switching does occur, we need to assume that business owners do not perfectly observe their productivity type (that is, $\sigma_\eta > 0$).

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

$$\begin{aligned} U(c,\ell) &= (c^{1-\psi}\ell^{\psi})^{1-\sigma}/(1-\sigma) + \xi \\ f_\kappa(h_\kappa, e) &= h_\kappa^{\vartheta} e^{1-\vartheta} \\ f_y(\kappa, h_y, k, n) &= \kappa^{\phi} k^{\alpha} (\omega h_y^{\rho} + (1-\omega) n^{\rho})^{\frac{\nu}{\rho}}, \end{aligned}$$

with values for parameters listed in Table 5.

There are four parameters related to preferences: ψ, ξ, σ, and β. Setting the weight on leisure, ψ, to 58 percent ensures that levels of aggregated business hours are consistent with U.S. totals. The parameter ξ captures the nonpecuniary benefit of running a business and is typically chosen to deliver a positive earnings differential between paid- and self-employment. As we have shown throughout, our data imply that there are significant *pecuniary* benefits and therefore we set this parameter to 0. The value for σ of 1.5 is standard in the literature. The value of $\beta = 0.96$ is consistent with U.S. real interest rates of roughly 4 percent.

Next, consider parameters of technologies. The most relevant for the income and growth profiles are the share of intangible capital in the production of goods and services, ϕ, the share of owner hours in the production of intangible capital, ϑ, and the elasticity between owner hours and external hours in the production of goods and services, ρ. The share ϕ affects the founders' incentives to invest time and resources in building the business. If the revenue share is small and external factors are used, then growth of the productive self-employed will be high relative to paid-employed or entrepreneurs that switch out of self employment early. Relatedly, the external resource requirement versus own time in intangible capital production governed by φ plays an important role for owner time use early in the career. The substitutability of owner and employee time ρ determines the scalability of the business. If they are highly substitutable, then the owner can create the intangible asset, say the customer base, and the employees work with it. To estimate these parameters, Bhandari and McGrattan (2021) used 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. These data informed the following choices: $\phi = 0.15$, $\vartheta = 0.408$, and $\rho = 0.5$. In terms of the other production shares, namely, $\alpha = 0.3$, $\omega = 0.425$, and $\nu = 0.55$, the estimates can be shown to be consistent with 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 $\delta_k = 0.041$ and $\delta_\kappa = 0.058$.

Prices are taken as given by the entrepreneurs when selling their goods or businesses and when hiring external labor and capital. As shown in Table 5, we set the interest rate to 4.1 percent, as noted above, and normalize the wage rate w to 1. The price of goods and services p, along with permanent values for the entrepreneurial productivity $\bar{z}_0$, determine average incomes at age 40 and are chosen to ensure incomes for the young entrepreneurs—both stayers and switchers—that are consistent with the IRS data. The price per unit of self-created capital, p_κ, if sold is set equal to 1.65 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.

The final parameters are those governing entrepreneurial and employee productivity. We need to choose initial conditions for the predicted mean μ_0 and variance σ_0^2 and the variance of the productivity shocks σ_η^2. We normalize the average of $\bar{z}_0$ equal to 0 and set the predicted mean to this value. The initial variance is chosen to generate realistic cross-sectional heterogeneity and the shock variance is chosen to generate sufficient delay before switching occurs. In Table 5, panel B, we report the transition matrix for employee productivity, which is relevant for entrepreneurial exit decisions. These estimates are consistent with the estimated wage processes of Low, Meghir, and Pistaferri (2010) for U.S. households in SIPP and PSID.

Given parameter estimates, we can now use our laboratory to simulate income and growth profiles for a large sample of entrepreneurs. In the simulations, we assume a potential working life 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. Initial financial and intangible assets are equal to 0 and there are transfers equal to 0.01 so that initial consumption is not 0. 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 is 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 19, we show the differenced income growth from the data (that is, differences in the two profiles in Figure 16) 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 humped growth profile, although predicted differences for the model are slightly higher. At the peak, the model predicts a growth differential of roughly \$6 thousand occurring around age 32 whereas we found it to be \$7.1 thousand around age 34 in our IRS sample.

To understand these results, it helps to consider the role of two key features: learning and investment. If we abstract from either, then we are unable to generate the patterns in Figure 19. 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 μ_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 σ_η^2 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 κ made by entrepreneurs. For our baseline parameters, we find that roughly 10 percent of available time is used initially to build κ. By age 35, investment is close to zero for the switchers but still 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, κ. In Figure 20, we report the predicted growth differential estimates as we vary ϕ 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, despite the fact that the life-cycle growth patterns differ. As Panel A of Figure 20 shows, the choice of ϕ can have a large effect on the differential growth between entrepreneurs that continue and those that exit. In the baseline parameterization, we set ϕ to 0.15. In the figure, we show growth differentials as we lower ϕ to 0.1 and even further to 0.05. The associated investments in each case are shown in Panel B of Figure 20. 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. How much they scale up and when depends on the specification of hours in production, for example, the share of owner time, ω, and its substitutability with external labor, ρ, as well as the intangible capital production, f_κ, since owner time is used for two activities. But varying these parameters does not change the overall message that our predictions for the growth differentials depends 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

This paper uses U.S. administrative tax dat a to characterize life-cycle income and growth profiles for self-employed individuals, comparing and contrasting them with paid employees. We find large differences once we classify individuals in terms of their attachment to work and their relative attachments to self-, paid-, or non-employment. The relatively attached self-employed have significantly higher average incomes over the life cycle and the most successful have persistent growth profiles. The data are compared to predictions of a theory of entrepreneurs that make firm-specific investments and learn about their productive abilities as they gain experience. The model generates comparable income and growth profiles and is used to study the contributions to income growth.

References

Alba-Ramirez, Alfonso. 1994. "Self-employment in the midst of unemployment: the case of Spain and the United States." *Applied Economics* 26 (3):189–204.

Bayard, Kimberly, U S Census Bureau, Emin Dinlersoz, University of Notre Dame, Timothy Dunne, University of Maryland, John Haltiwanger, Javier Miranda, and John Stevens. 2018. "Early-Stage Business Formation: An Analysis of Applications for Employer Identification Numbers." *NBER Working Paper* 2018 (015).

Buera, Francisco J. 2009. "A dynamic model of entrepreneurship with borrowing constraints: theory and evidence." *Annals of Finance* 5 (3-4):443–464.

Cagetti, Marco and Mariacristina DeNardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy* 114 (5):835–870.

Catherine, Sylvain. 2022. "Keeping options open: What motivates entrepreneurs?" *Journal of Financial Economics* 144 (1):1–21.

DeBacker, Jason, Vasia Panousi, and Shanthi Ramnath. 2022. "A Risky Venture: Income Dynamics Among Pass-Through Business Owners." *American Economic Journal: Macroeconomics* .

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28 (3):3–24.

Evans, David S and Boyan Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy* 97 (4):808–827.

Evans, David S. and Linda S. Leighton. 1989. "Some Empirical Aspects of Entrepreneurship." *American Economic Review* 79 (3):519–535.

Fairlie, Robert W. 2005. "Self-employment, entrepreneurship, and the NLSY79." *Monthly Labor Review* :40–47.

Garin, Andrew, Emilie Jackson, and Dmitri Koustas. 2022. "New Gig Work or Changes in Reporting? Understanding Self-Employment Trends in Tax Data." *BFI Working Paper* .

Hall, Robert E and Susan E Woodward. 2010. "The Burden of the Nondiversifiable Risk of Entrepreneurship." *American Economic Review* 100 (3):1163–1194.

Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Journal of Political Economy* 108 (3):604–631.

Hurst, Erik and Annamaria Lusardi. 2004. "Liquidity Constraints, Household Wealth, and Entrepreneurship." *Journal of Political Economy* 112 (2):319–347.

Hurst, Erik and Benjamin Wild Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* 2011 (2):73–118.

Kartashova, Katya. 2014. "Private Equity Premium Puzzle Revisited." *American Economic Review* 104 (10):3297–3334.

Lazear, Edward P and Robert L Moore. 1984. "Incentives, Productivity, and Labor Contracts." *Quarterly Journal of Economics* 99 (2):275.

Levine, Ross and Yona Rubinstein. 2016. "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?*." *Quarterly Journal of Economics* 132 (2):963–1018.

Lise, Jeremy and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *American Economic Review* 110 (8):2328–76.

Moskowitz, Tobias J and Annette Vissing-Jorgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92 (4):745–778.

Parker, Simon C. 2018. *The Economics of Entrepreneurship*. Cambridge University Press, 2 ed.

Quadrini, Vincenzo. 1999. "The Importance of Entrepreneurship For Wealth Concentration and Mobility." *Review of Income and Wealth* 45 (1):1–19.

Rissman, Ellen. 2003. "Self-Employment as an Alternative to Unemployment." *Federal Reserve Bank of Chicago Working Paper Series* .

———. 2007. "Labor Market Transitions and Self-Employment." *Federal Reserve Bank of Chicago Working Papers* .

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics* 134 (4):1675–1745.

Table 1: Main Sample Summary Statistics

Statistic	Total Sample	Attached PE	Attached SE	Almost PE	Almost SE	Mostly Switching	Any NE
Individuals (Mill.)	65.0	35.4	1.9	0.5	0.3	2.2	24.8
Shares (%)							
Counts	100.0	54.4	2.9	0.7	0.4	3.4	38.1
Total income	100.0	66.8	8.4	1.3	1.1	6.5	15.9
PE income	100.0	77.2	1.0	1.4	0.4	4.6	15.4
SE income	100.0	4.1	52.5	0.8	5.5	18.3	18.8
Incomes (2012 \$, Thous.)							
Mean PE + SE income	53.5	65.6	154.4	92.9	148.7	103.3	22.2
Income, 10^{th} pctl	6.7	23.7	17.3	22.1	21.4	17.9	2.1
25^{th}	17.6	33.8	31.4	33.5	38.8	29.5	6.4
50^{th}	35.9	49.7	66.8	54.0	73.4	53.2	14.3
75^{th}	61.3	73.8	153.2	92.3	147.4	103.8	26.8
90^{th}	99.7	110.9	334.8	162.9	306.1	206.2	44.8
Mean PE	45.8	65.0	16.4	84.9	46.4	62.0	18.5
Income, 10^{th} pctl	3.4	23.6	0.0	21.1	4.5	10.7	0.7
25^{th}	13.9	33.8	0.0	31.8	8.3	18.4	4.2
50^{th}	32.8	49.5	2.0	50.1	17.6	33.0	11.9
75^{th}	57.1	73.3	10.4	83.4	41.7	61.0	23.9
90^{th}	90.1	109.6	32.8	144.7	91.7	113.3	40.0
Mean SE	7.6	0.6	138.0	8.1	102.4	41.3	3.8
Income, 10^{th} pctl	−0.1	−0.1	14.8	−4.6	10.0	0.5	0.0
25^{th}	0.0	0.0	27.1	0.1	20.6	6.0	0.0
50^{th}	0.0	0.0	58.4	3.2	46.2	15.4	0.0
75^{th}	0.5	0.0	136.2	9.2	102.2	39.8	1.0
90^{th}	9.0	1.0	301.4	21.9	219.8	94.7	7.6

Table 1: Main Sample Summary Statistics (Cont.)

Statistic	Total Sample	Attached PE	Attached SE	Almost PE	Almost SE	Mostly Switching	Any NE
Incomes (2012 $, Thous.)							
Spousal wages, Mean	39.1	35.7	31.0	31.5	36.1	34.3	45.1
Median	25.7	27.1	14.9	18.2	20.1	19.2	25.4
UI income, Mean	0.4	0.3	0.0	0.5	0.1	0.3	0.5
Median	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gig work, Mean	3.0	1.9	7.9	6.2	6.5	7.3	3.5
Median	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Asset income, Mean	7.6	4.7	36.8	16.6	48.6	20.8	7.7
Median	0.0	0.1	0.8	0.1	2.1	0.2	0.0
Income tax rates at age 40 (%)	30.3	29.1	35.7	37.0	35.7	34.9	32.4
Skills (%)							
Educated	52.1	59.6	61.6	66.2	71.8	63.7	39.3
Cognitive	47.2	52.7	59.9	59.2	63.7	58.6	37.0
Interpersonal	56.1	63.1	58.7	65.4	68.8	62.8	45.0
Manual	32.3	31.1	32.7	31.1	26.6	30.1	34.2
Industries (%)							
Agriculture	1.1	0.9	1.7	1.3	1.8	1.3	1.4
Mining	0.4	0.5	1.5	0.7	0.6	0.6	0.3
Utilities	0.2	0.3	0.1	0.2	0.1	0.1	0.1
Construction	7.3	5.4	17.2	10.3	14.9	12.7	8.8
Manufacturing	12.0	14.8	3.6	10.7	4.7	6.3	9.1
Wholesale trade	3.8	4.1	3.5	4.3	4.1	3.9	3.3
Retail trade	9.0	8.1	7.9	8.4	9.4	8.4	10.4
Transportation	3.4	3.1	4.8	4.8	4.1	5.2	3.6
Information	1.8	2.0	1.0	2.6	1.4	1.7	1.5
Finance	3.6	4.0	4.6	4.4	4.4	4.3	2.9
Real estate	2.4	1.8	4.7	3.5	4.8	4.1	3.0

Table 1: Main Sample Summary Statistics (Cont.)

Statistic	Total Sample	Attached PE	Attached SE	Almost PE	Almost SE	Mostly Switching	Any NE
Industries (%)							
Professional	9.9	9.5	16.1	15.8	17.0	17.2	9.2
Management	0.6	0.9	0.1	0.6	0.2	0.2	0.3
Administration	4.8	3.8	4.4	4.5	4.0	4.3	6.5
Education	0.6	0.5	0.4	0.7	0.5	0.7	0.7
Health care	6.6	5.6	10.6	6.4	10.2	9.5	7.5
Arts	1.1	0.8	1.8	1.5	2.3	1.7	1.3
Accommodation	4.2	3.1	3.4	3.9	4.8	4.1	5.9
Other services	4.0	2.1	11.1	4.2	8.9	7.8	5.7
Other NAICS	12.7	16.5	0.5	10.4	0.6	4.3	9.2
Missing NAICS	10.3	12.2	1.9	0.7	1.2	1.4	9.2
Demographics							
Male (%)	50.7	53.4	82.4	75.1	79.4	75.1	41.5
Mostly married (%)	67.6	70.3	79.1	73.6	82.0	75.4	61.9
Has children (%)	82.5	82.7	84.9	86.0	87.0	85.8	81.6
Number of children (mean)	2.2	2.2	2.2	2.4	2.3	2.3	2.3
Birth year (median)	1963	1963	1960	1964	1961	1963	1964

Note: PE=paid-employment, SE=self-employment, NE=non-employment. See Sections 3 and 4 for details on the samples and subgroups. Percentiles are an average of the observations around the relevant percentile (in order to avoid revealing an individuals' information).

Table 2: Largest Contributors to Attached-Employee Growth Gap

		Characteristics					
Cumulative Share	NAICS	Male	Married	Educated	Cognitive	Interpersonal	Manual
15.4	62	√	√	√	√	√	
26.7	54	√	√	√		√	
33.1	54	√	√	√	√	√	
39.4	52	√	√	√		√	
44.9	62	√	√	√	√	√	√
49.3	44	√	√	√	√	√	
53.5	23	√	√	√	√	√	

Note: The data underlying this table is the sample of attached self- and paid-employees. Shares are each groups contribution to the self- and paid-employment growth gap between ages 30 and 39.

Table 3: Transition Probabilities For Attached Self- and Paid-Employment Income (thousands of 2012 dollars)

Attached Self-Employed

Income In Period t	Income In Period $t-1$: <7	7-11	11-17	17-27	27-43	43-68	68-106	106-168	168-264	264-415	>415
<7	49	19	11	9	7	5	4	3	2	2	3
7-11	6	29	13	5	2						
11-17	6	24	37	16	5	2					
17-27	7	14	23	39	17	6	2				
27-43	7	7	9	21	40	19	6	2			
43-68	6	4	4	7	19	40	21	6	2		
68-106	5	2	2	3	6	19	41	21	6	2	
106-168	4				2	5	18	44	21	6	2
168-264	3						4	17	45	21	4
264-415	2							4	16	47	12
>415	3								4	19	77
Distribution	4	3	7	11	14	14	13	11	9	6	8

Attached Paid-Employed

Income In Period t	Income In Period $t-1$: <7	7-11	11-17	17-27	27-43	43-68	68-106	106-168	168-264	264-415	>415
<7	26	11	4								
7-11	20	33	11	3							
11-17	19	26	45	10	2						
17-27	18	18	27	61	11	2					
27-43	11	9	10	21	72	12					
43-68	4	3	3	3	13	76	15	2			
68-106						9	75	18	3		
106-168							8	70	21	5	2
168-264								8	61	22	6
264-415									12	55	16
>415									2	16	74
Distribution	1	2	4	11	24	29	19	8	2	1	1

Note: The sample underlying these transition probabilities is the entire set of attached self- and paid-employed individuals in our sample. The transition matricies show the probabilities of an individual in an income bin at time $t-1$ remaining in their current income bin or moving to a different income bin at time t. The distribution line displays the distribution of person-years between income bins.

Table 4: Income Shares Held by Each Employment Group

Total Income

Percentile Group	Total Sample	Attached PE	Attached SE	Almost PE	Almost SE	Mostly Switching	Any NE
$< 10^{th}$	0.8	0.1	-0.1	0.0	0.0	0.0	0.8
10^{th} to 25^{th}	4.4	1.6	0.1	0.0	0.0	0.1	2.6
25^{th} to 75^{th}	36.8	25.9	1.1	0.4	0.2	1.4	7.9
75^{th} to 90^{th}	21.8	17.0	1.2	0.3	0.2	1.2	1.9
$< 90^{th}$	36.2	22.1	6.1	0.7	0.8	3.8	2.7

Paid-Employment Income

Percentile Group	Total Sample	Attached PE	Attached SE	Almost PE	Almost SE	Mostly Switching	Any NE
$< 10^{th}$	1.1	0.1	0.0	0.0	0.0	0.0	1.0
10^{th} to 25^{th}	4.7	1.9	0.0	0.0	0.0	0.1	2.7
25^{th} to 75^{th}	39.9	30.1	0.1	0.4	0.1	1.1	8.0
75^{th} to 90^{th}	22.8	19.7	0.2	0.3	0.1	0.9	1.8
$< 90^{th}$	31.5	25.4	0.7	0.7	0.3	2.5	1.9

Self-Employment Income

Percentile Group	Total Sample	Attached PE	Attached SE	Almost PE	Almost SE	Mostly Switching	Any NE
$< 10^{th}$	-1.5	-0.1	-0.6	0.0	-0.1	-0.3	-0.3
10^{th} to 25^{th}	3.0	0.0	0.5	0.0	0.0	0.2	2.3
25^{th} to 75^{th}	18.6	0.8	6.9	0.2	0.7	3.3	6.7
75^{th} to 90^{th}	15.8	1.04	7.7	0.2	0.9	3.4	2.7
$< 90^{th}$	64.1	2.4	38.1	0.5	4.0	11.8	7.4

Note: The data underlying this figure is the entire baseline sample. The income shares are the shares of the total corresponding income type over the length of the panel. PE=paid-employment, SE=self-employment

Table 5. Baseline Model Parameters

A. Preferences, Technologies, Prices, and Entrepreneurial Productivity

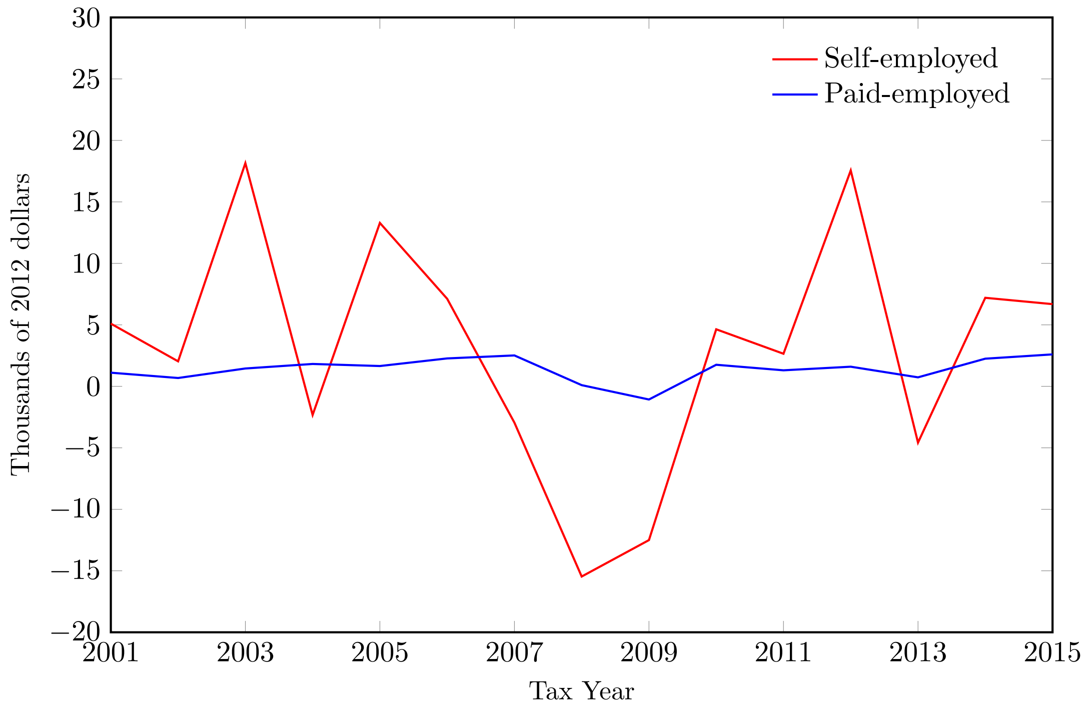
Parameter	Expression	Value
Preferences		
Leisure Weight	ψ	0.58
Love of business parameter	ξ	0.0
Intertemporal elasticity inverse	σ	1.5
Discount factor	β	0.96
Technologies		
Owner hours share, intangible production (%)	ϑ	40.8
Hours substitution parameter, goods production	ρ	0.5
Intangible capital share, goods production (%)	ϕ	15.0
Fixed asset share, goods production (%)	α	30.0
Owner hours share, goods production (%)	ω	42.5
Intangible capital depreciation (%)	δ_κ	5.8
Fixed asset depreciation (%)	δ_k	4.1
Prices		
Interest rate (%)	r	4.1
Hired labor	w	1.00
Goods and services	p	1.50
Intangible capital	p_k	1.65
Entrepreneurial Productivity		
Initial predicted mean	μ_0	0
Initial predicted variance	σ_0^2	0.005
Idiosyncratic shock variance	σ_η^2	0.008

B. Employee Productivity (ϵ_j) and Transition Probabilities

	Productivity at $j+1$				
Productivity at j	0.509	0.713	1.000	1.400	1.970
0.509	0.424	0.549	0.027	0.000	0.000
0.713	0.046	0.621	0.327	0.005	0.000
1.000	0.001	0.145	0.709	0.145	0.001
1.400	0.000	0.005	0.327	0.621	0.046
1.970	0.000	0.000	0.027	0.549	0.424

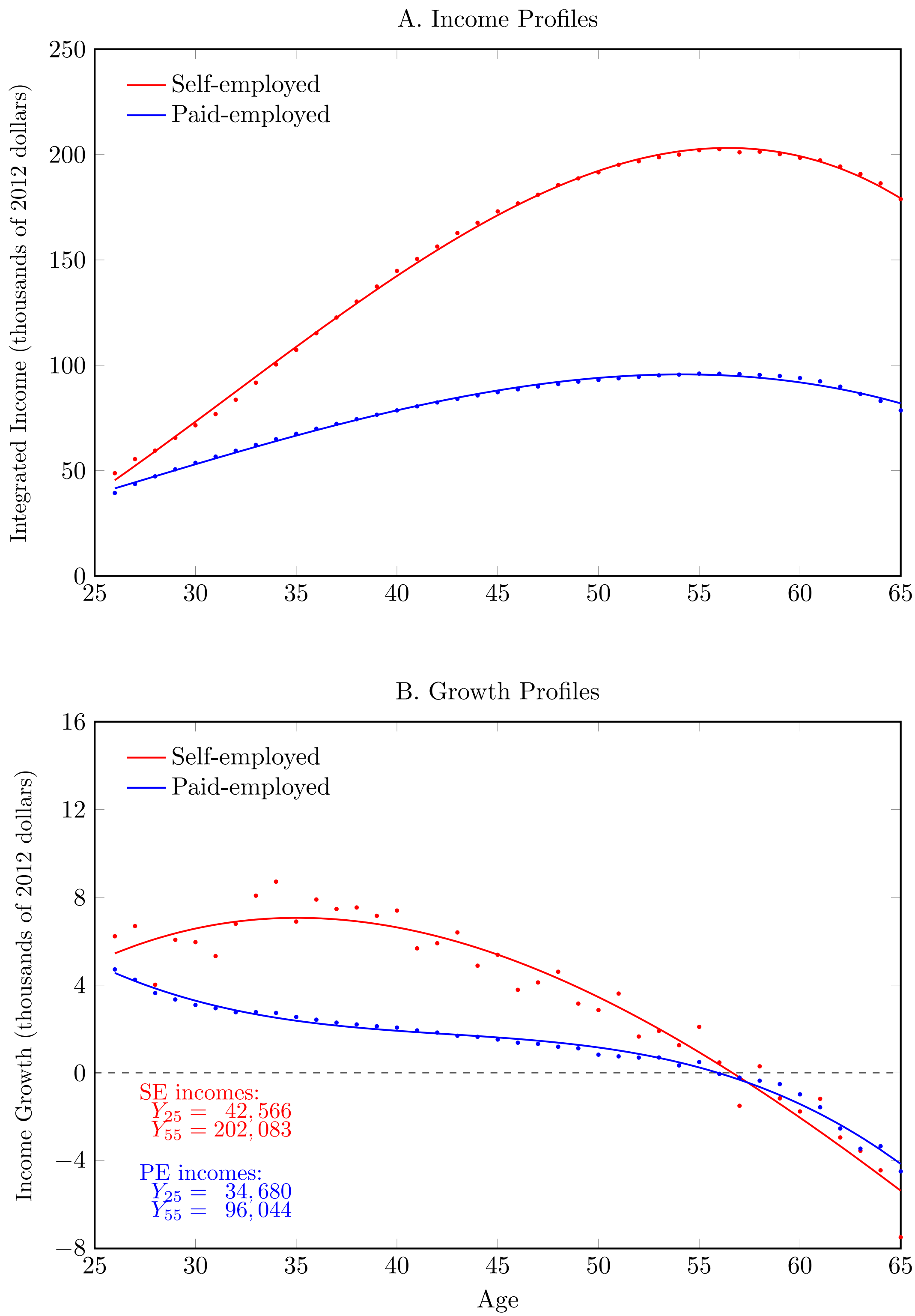
Note: See Section 5 for functional form assumptions, data sources, and details on parameter estimation.

Figure 1: Estimated Time Effects $\beta_{g,t}$ of Income



Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figure shows the time effects $\beta_{g,t}$.

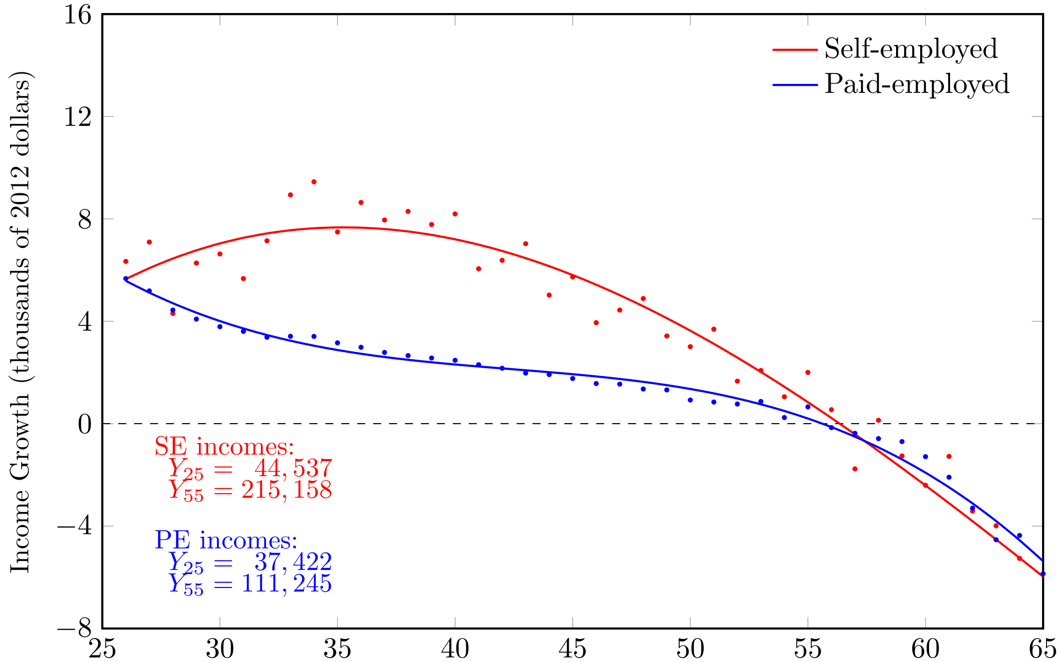
Figure 2: Income and Growth Profiles



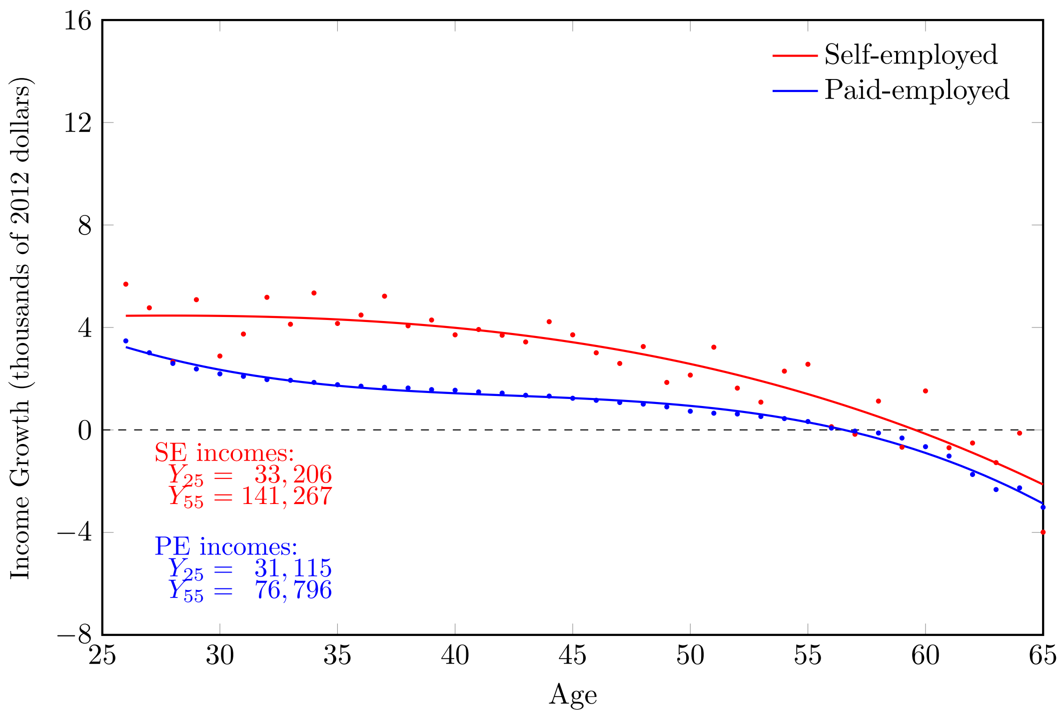
Note: The sample underlying these graphs are the attached self- and paid-employed groups. Panel A shows the integrated age effects, $Y_g(a)$, and Panel B shows the age effects, $\{\bar{\gamma}_g^a\}$.

Figure 3: Growth Profiles by Gender

A. Men

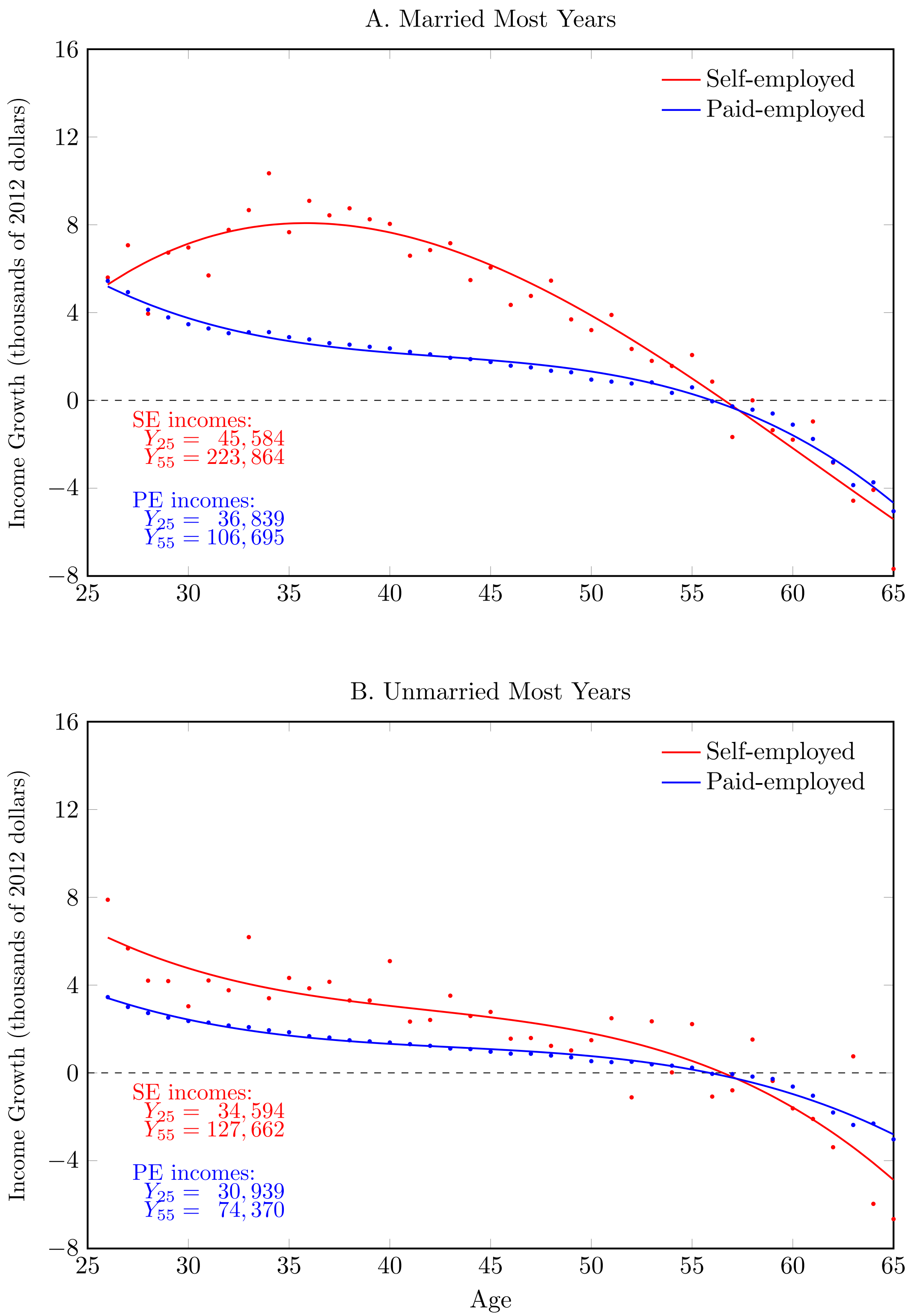


B. Women



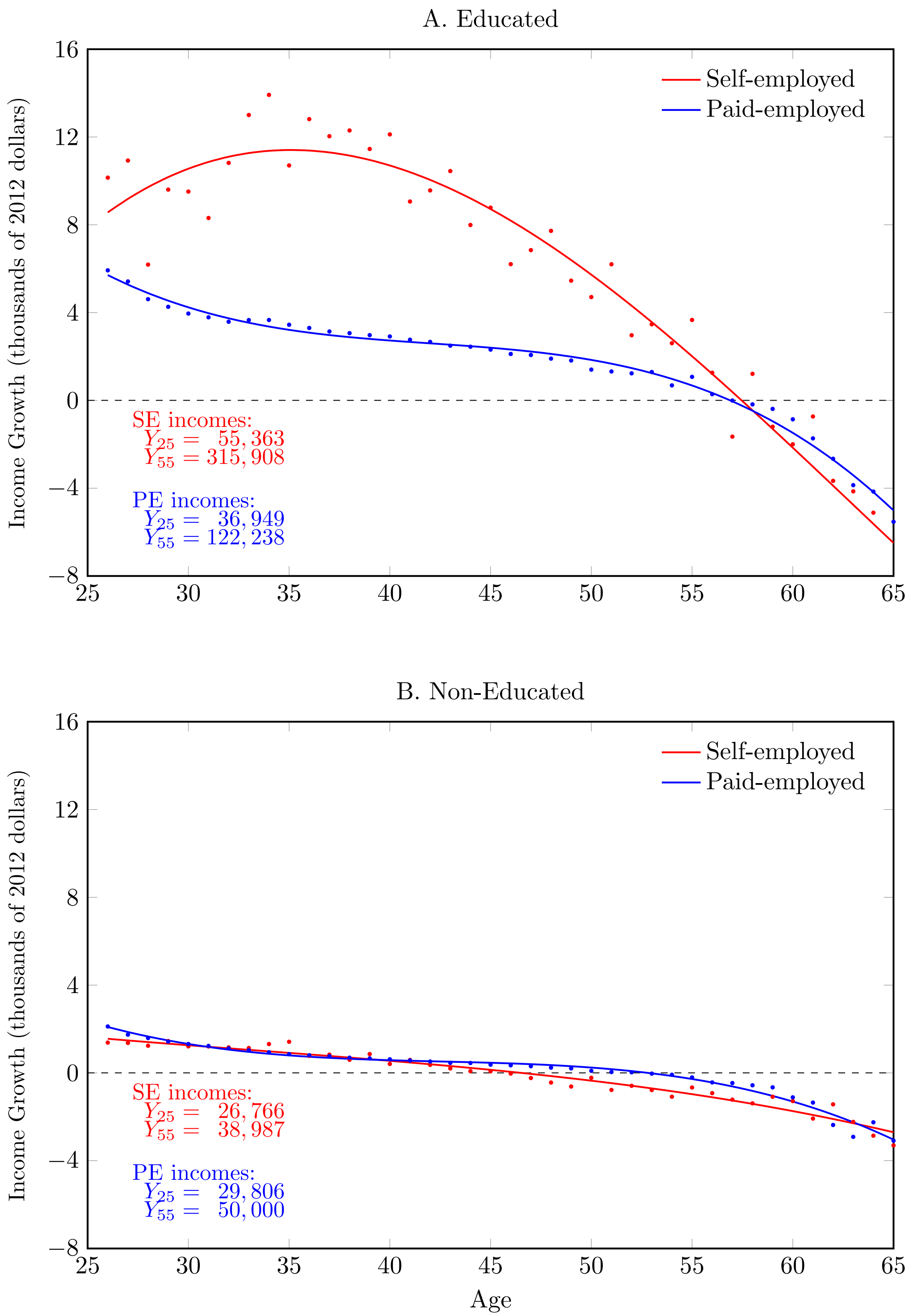
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for men (Panel A) and for women (Panel B).

Figure 4: Growth Profiles by Marital Status



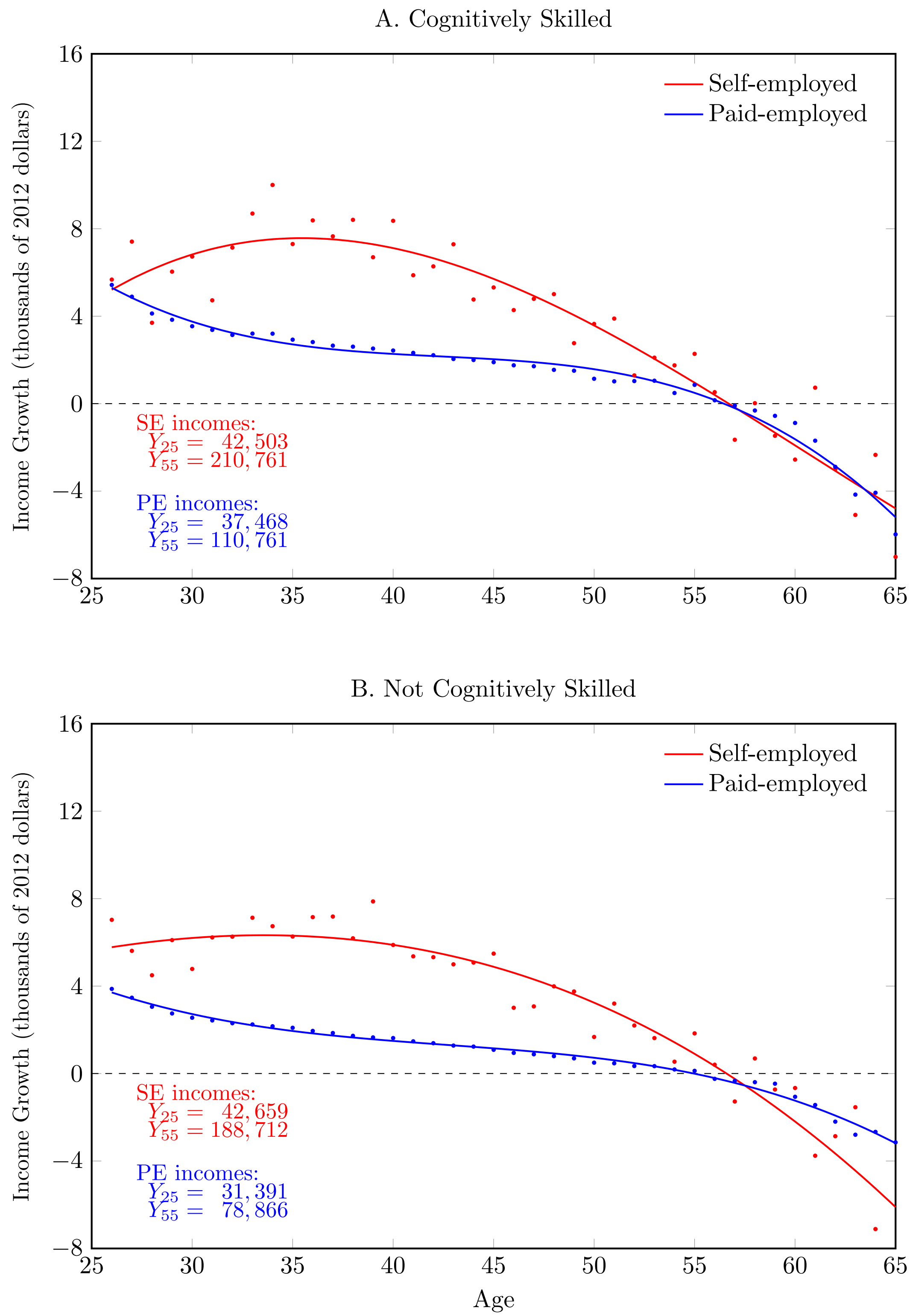
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for individuals who were married (Panel A) and who were unmarried (Panel B) during most years of the sample.

Figure 5: Growth Profiles by Education



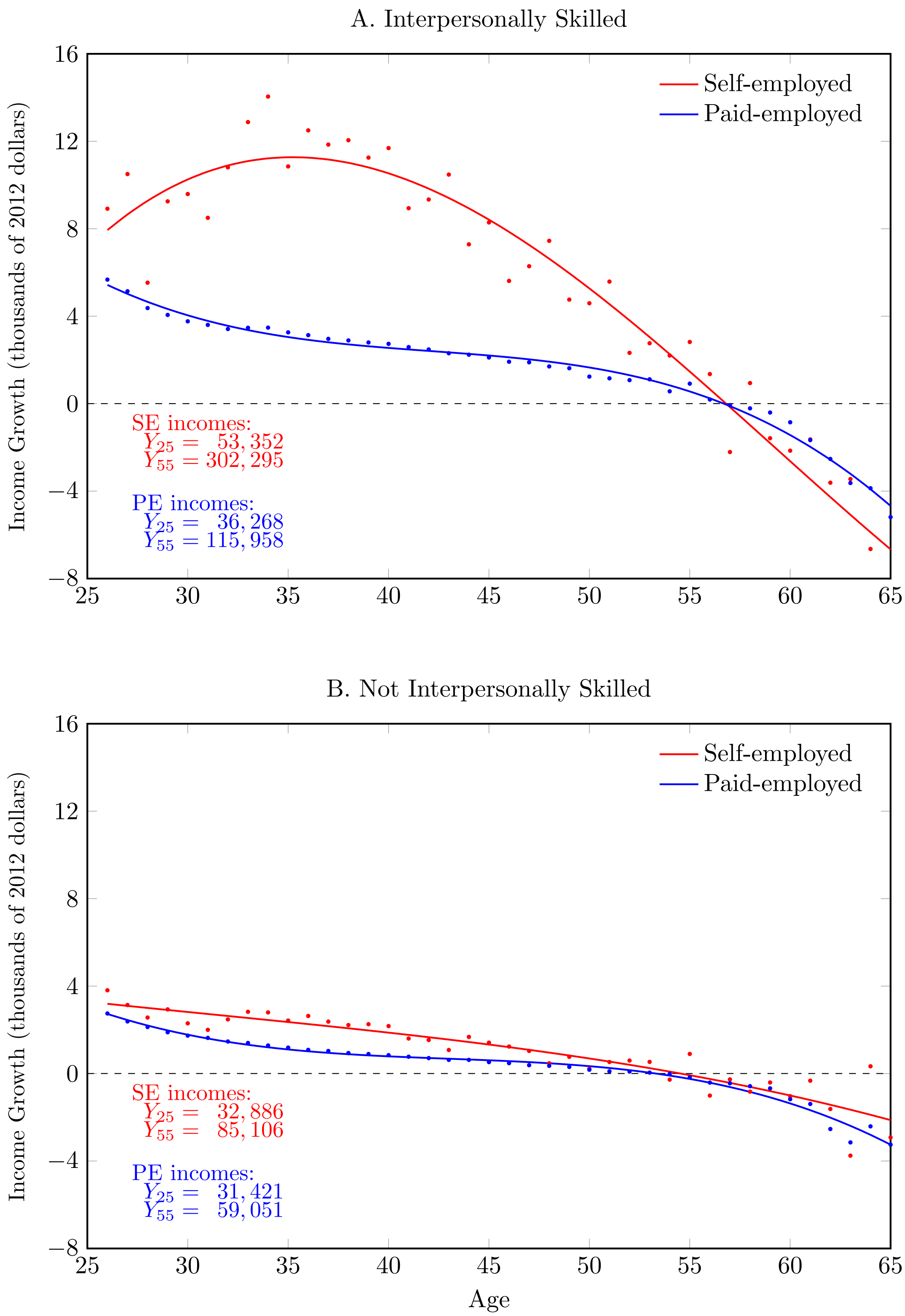
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for educated individuals (Panel A) and for non-educated individuals (Panel B).

Figure 6: Growth Profiles by Cognitive Skill



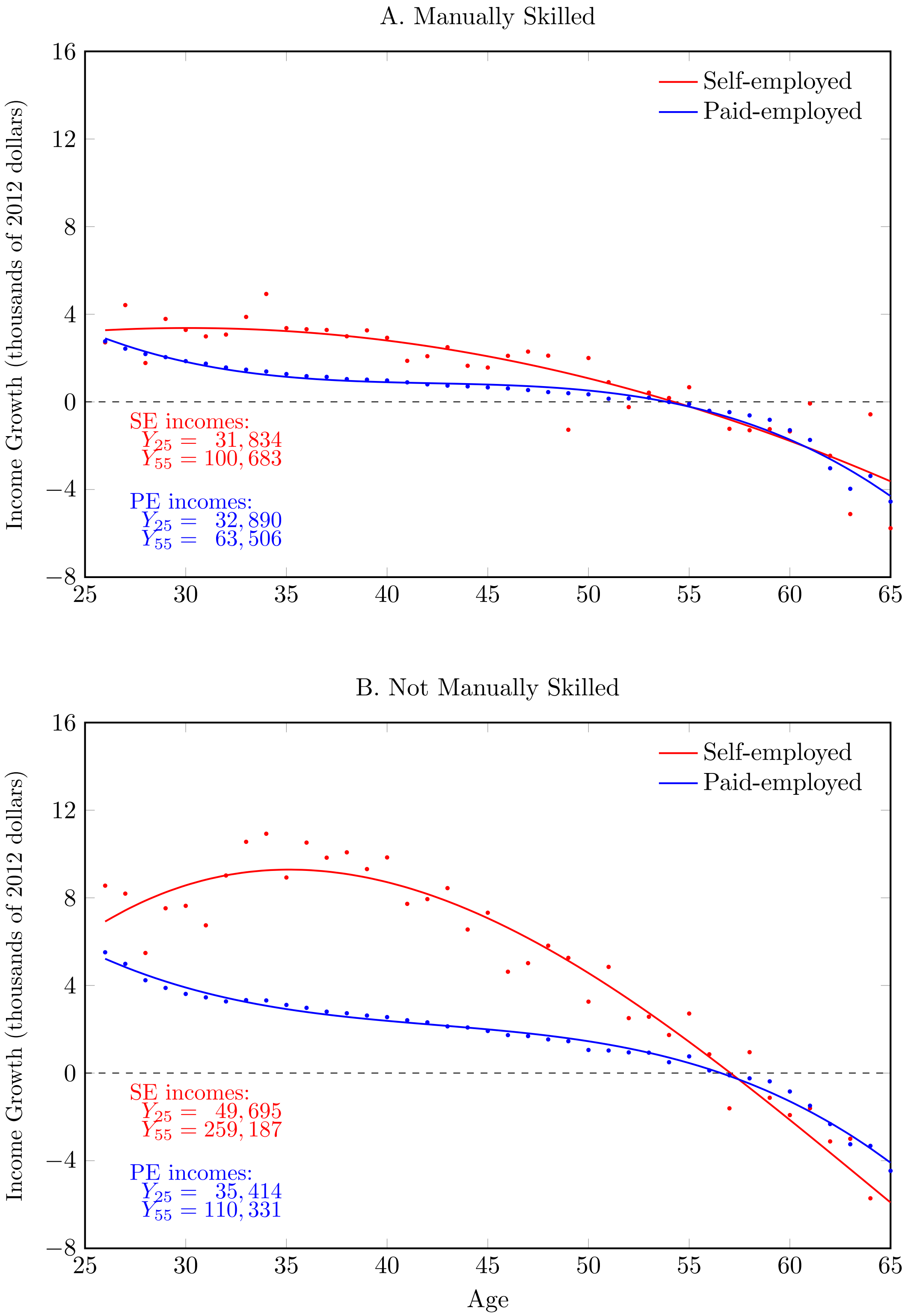
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for cognitively skilled individuals (Panel A) and for not cognitively skilled individuals (Panel B).

Figure 7: Growth Profiles by Interpersonal Skill



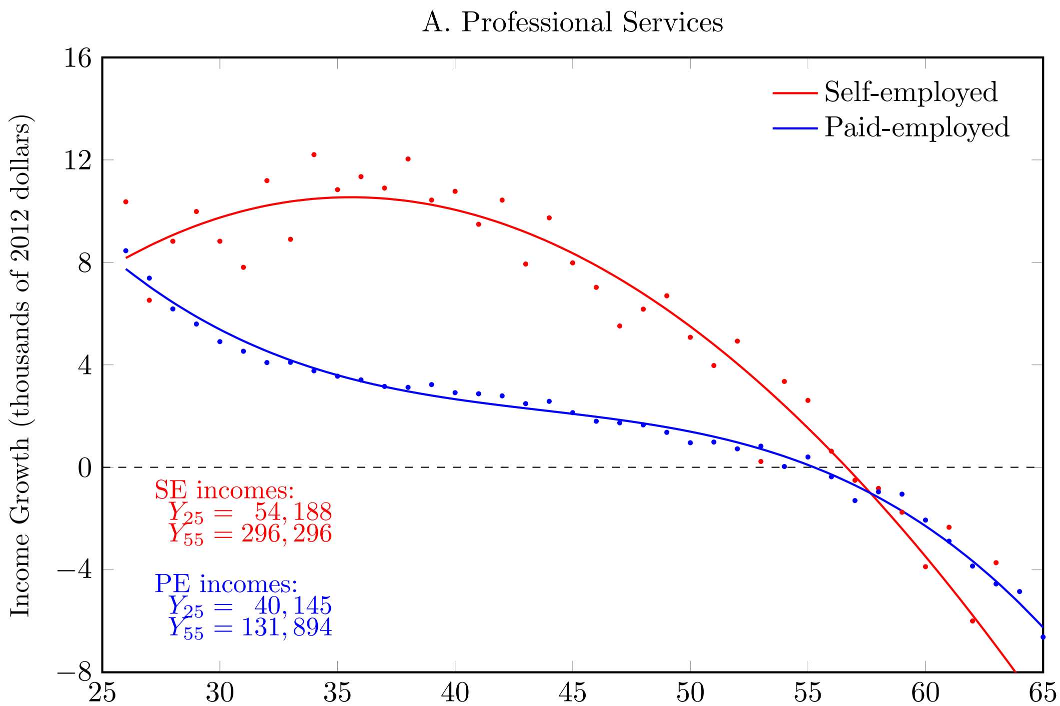
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for interpersonally skilled individuals (Panel A) and for not interpersonally skilled individuals (Panel B).

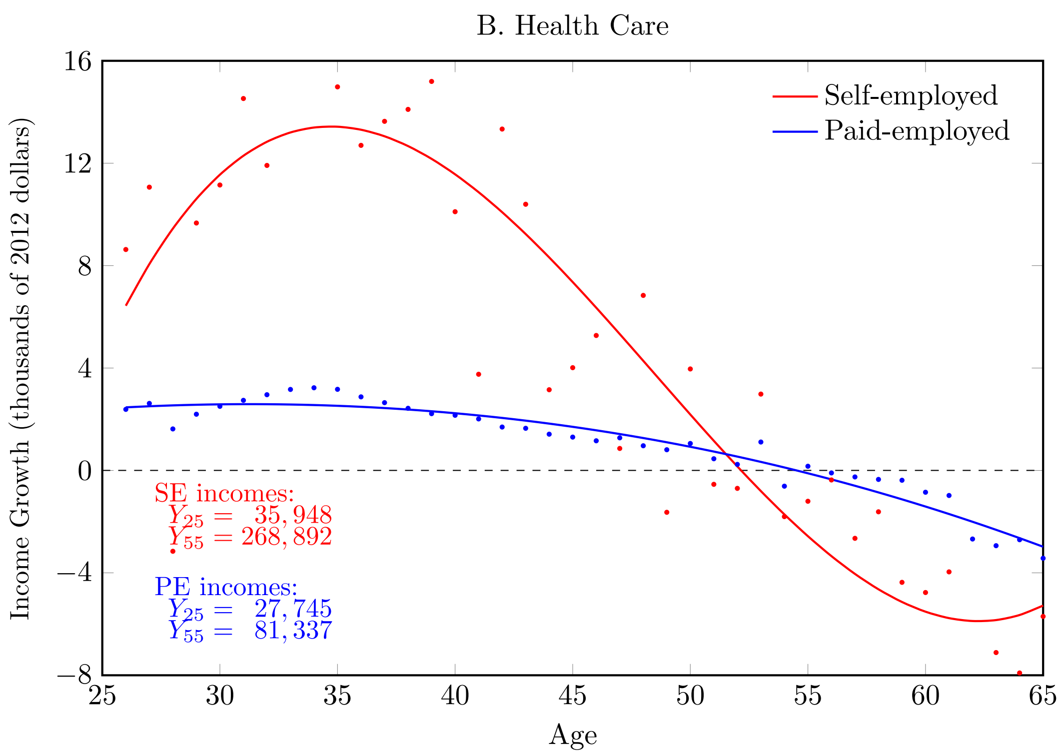
Figure 8: Growth Profiles by Manual Skill



Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for manually skilled individuals (Panel A) and for not manually skilled individuals (Panel B).

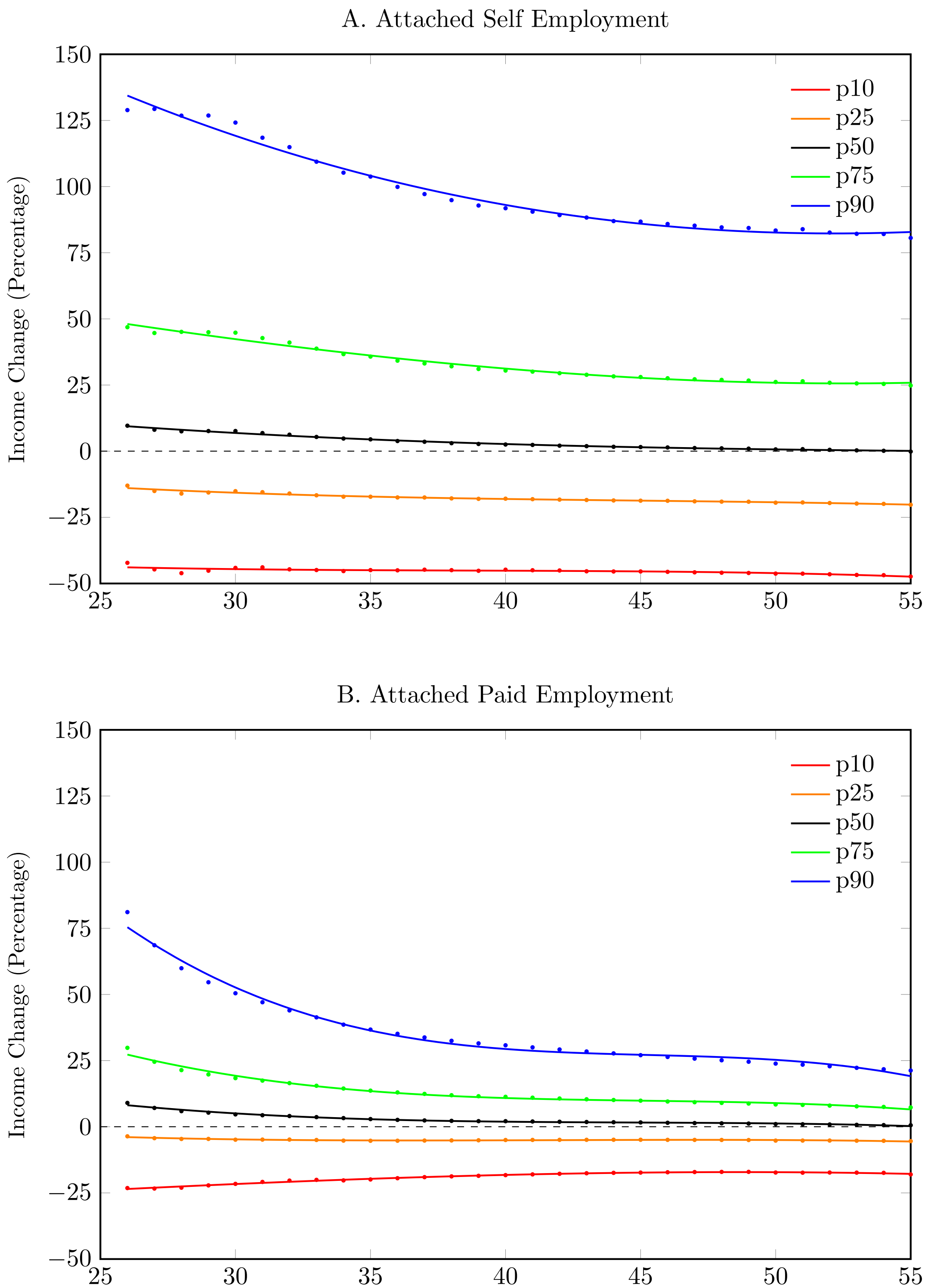
Figure 9: Growth Profiles by Industry





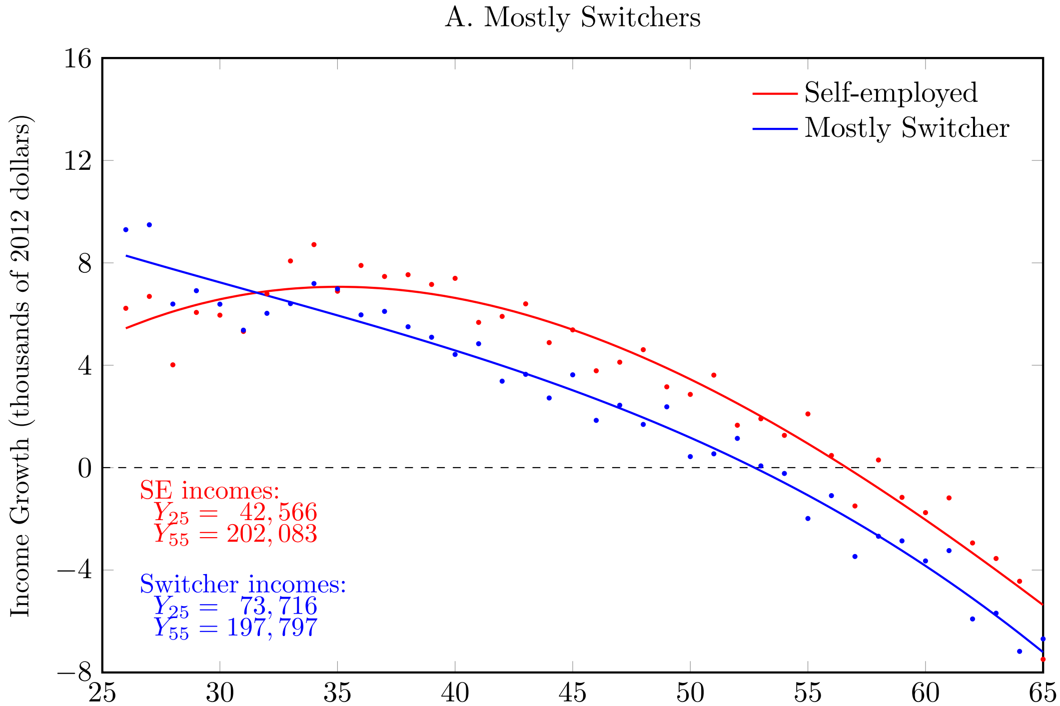
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figures show the age effects, $\{\bar{\gamma}^a_g\}$, for individuals in the professional services industry (Panel A) and in the health care industry (Panel B).

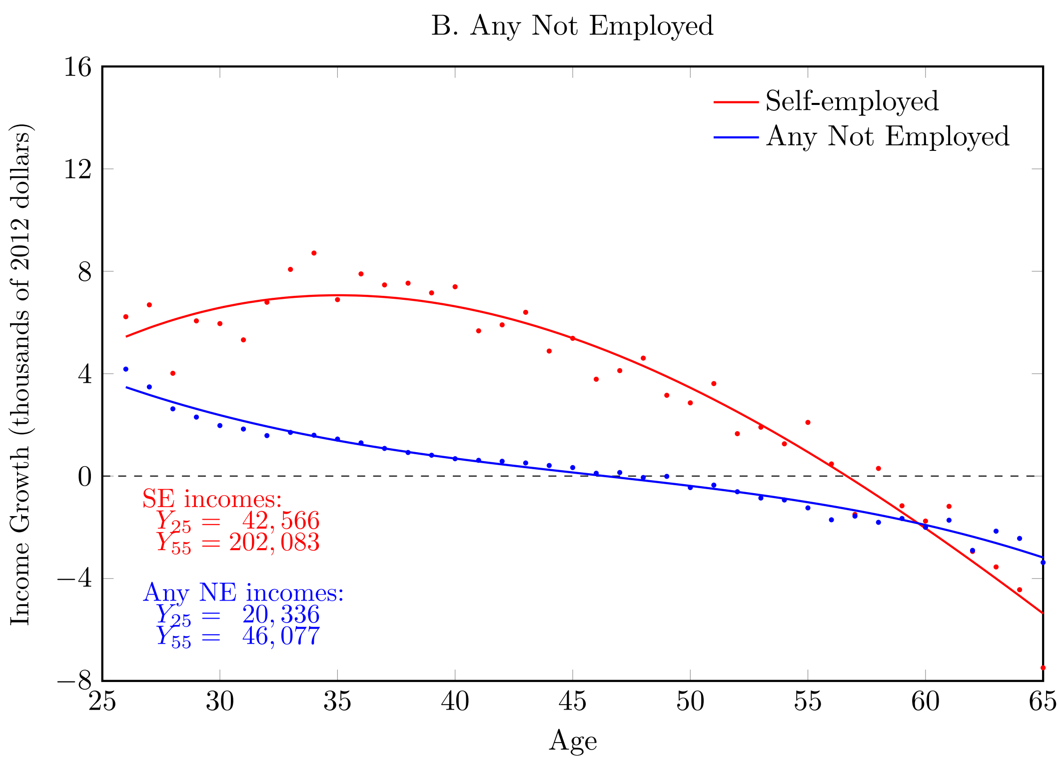
Figure 10: Age-Over-Age Growth



Note: The sample underlying these graphs are the attached self- and paid-employed groups. For each individual, we compute the age-over-age percentage change in income and plot selected percentiles of these changes. Panel A displays the percentiles for the attached self-employed group and Panel B displays the same for the attached paid-employed.

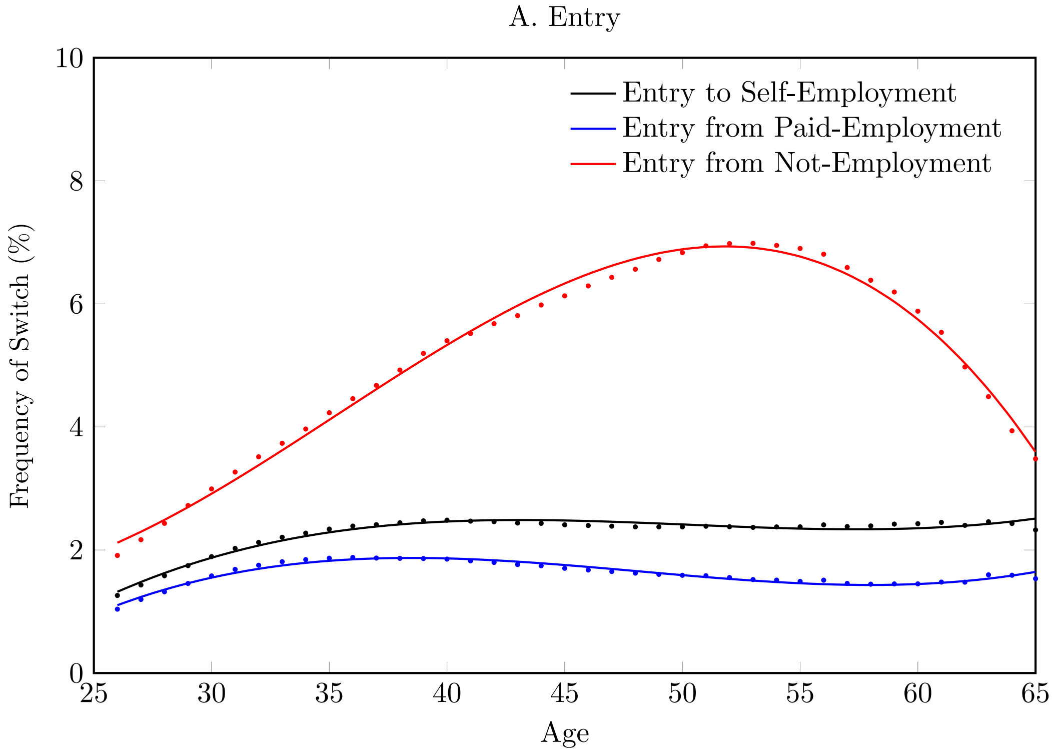
Figure 11: Growth of Mostly Switchers and Any Non-Employment

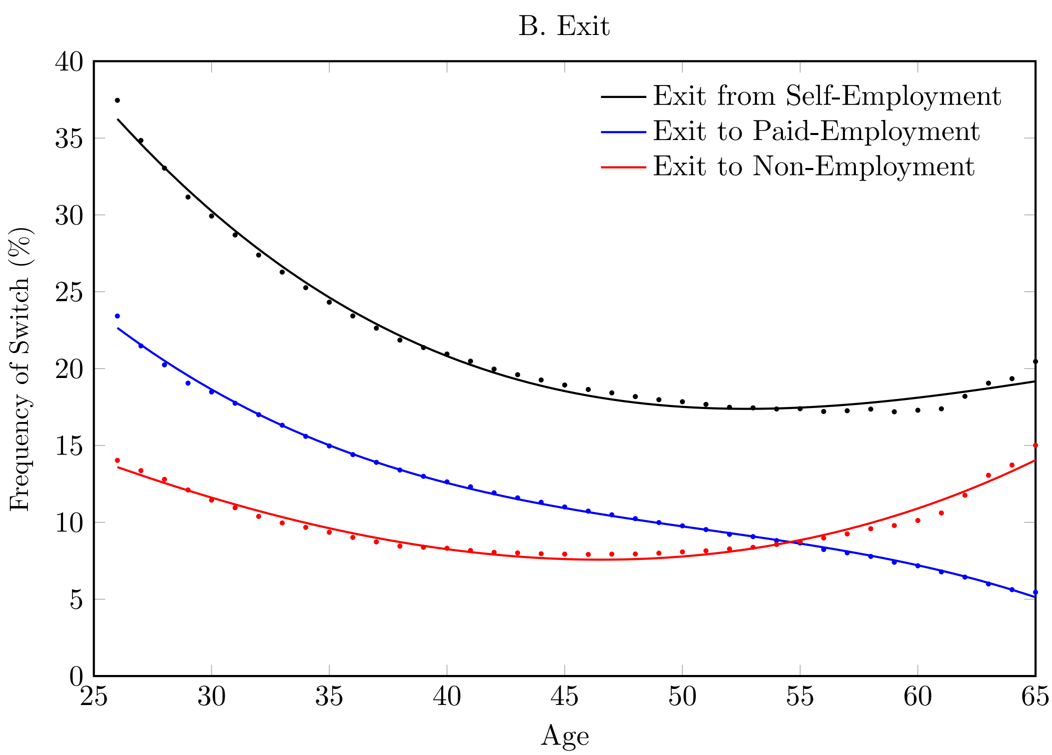




Note: The sample underlying these graphs are the switchers and any non-employed. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for individuals who are switchers (Panel A) and any non-employed (Panel B).

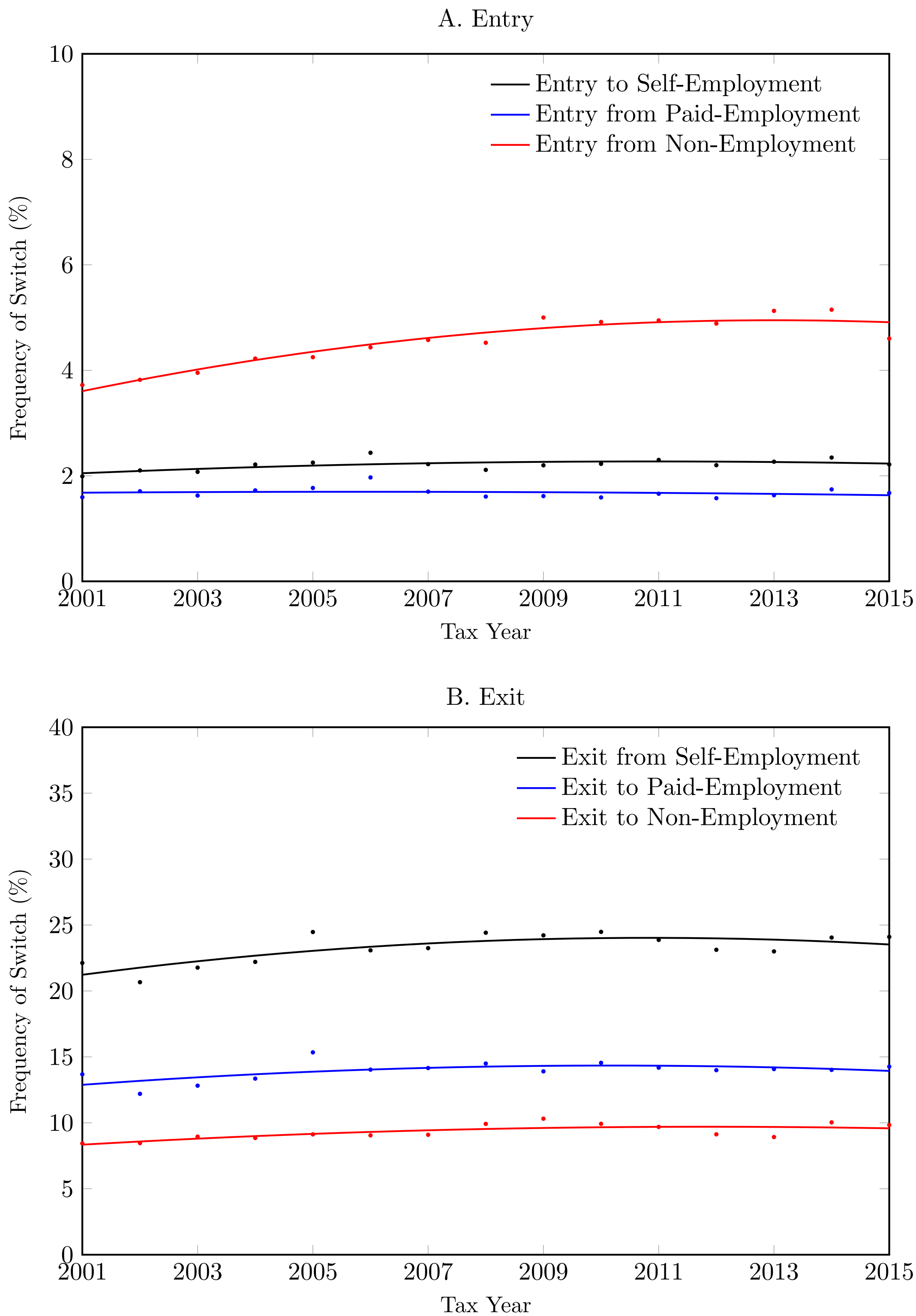
Figure 12: Self-Employment Switching Rate By Age





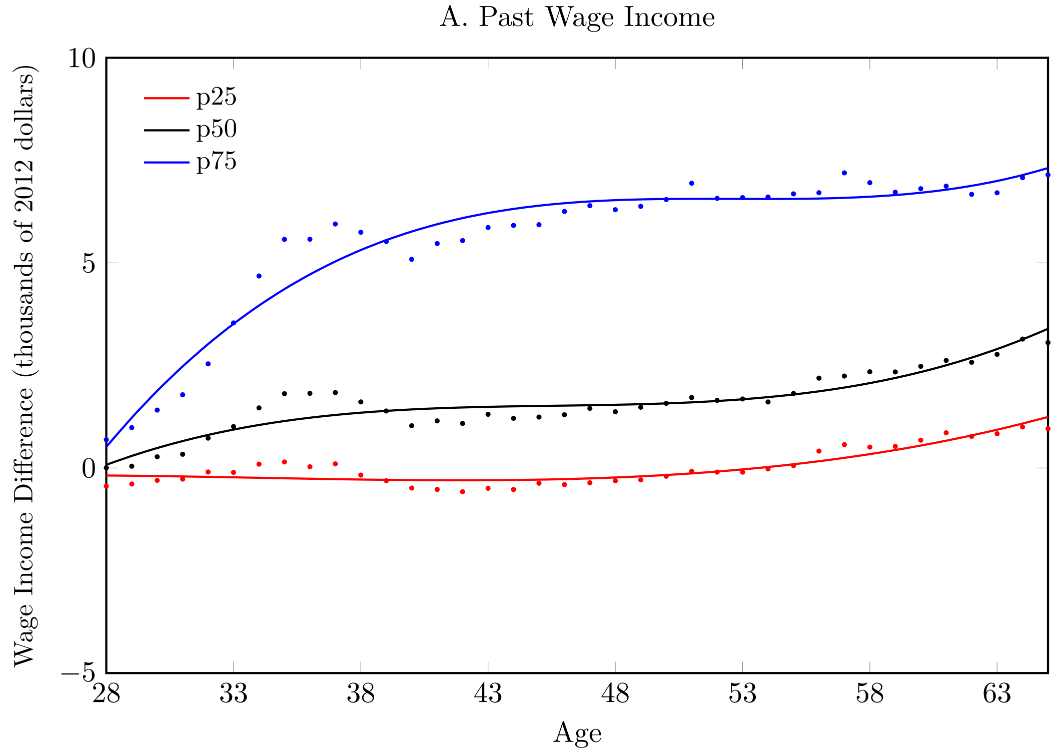
Note: The underlying sample is all individuals in our baseline sample. Panel A displays the entry rate to self-employment by age and Panel B displays the exit rate from self-employment by age.

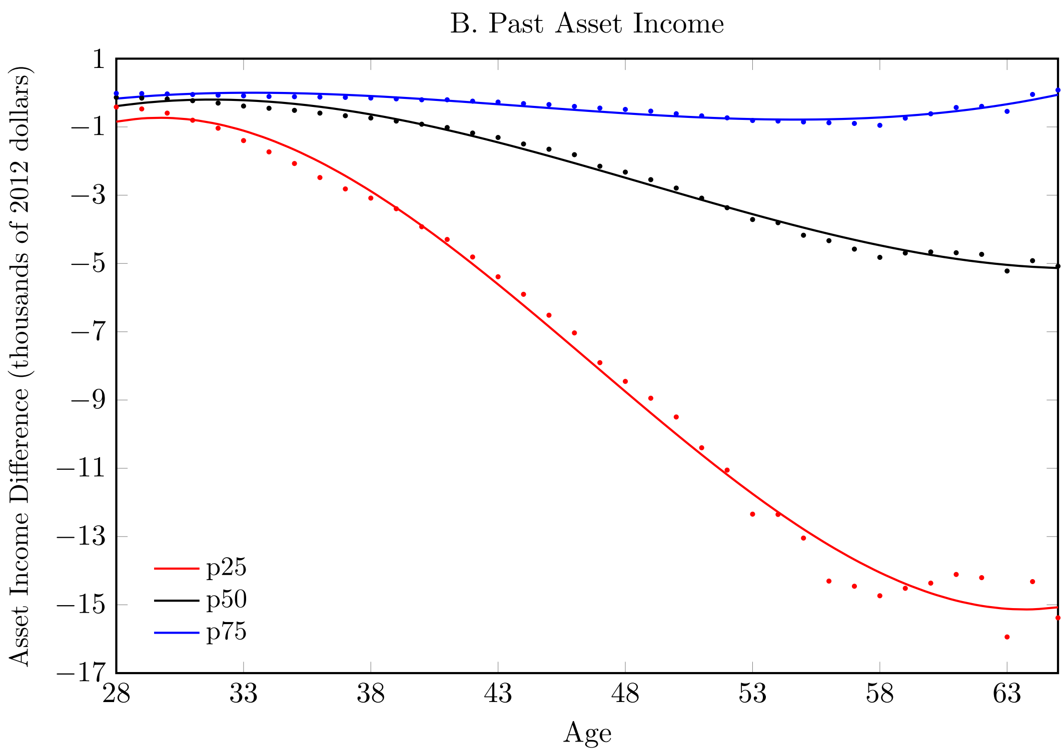
Figure 13: Self-Employment Switching Rate By Year



Note: The underlying sample is all individuals in our baseline sample. Panel A displays the entry rate to self-employment by year and Panel B displays the exit rate from self-employment by year. Each of these time series have been corrected for our panel's aging population (see Section 4 for details).

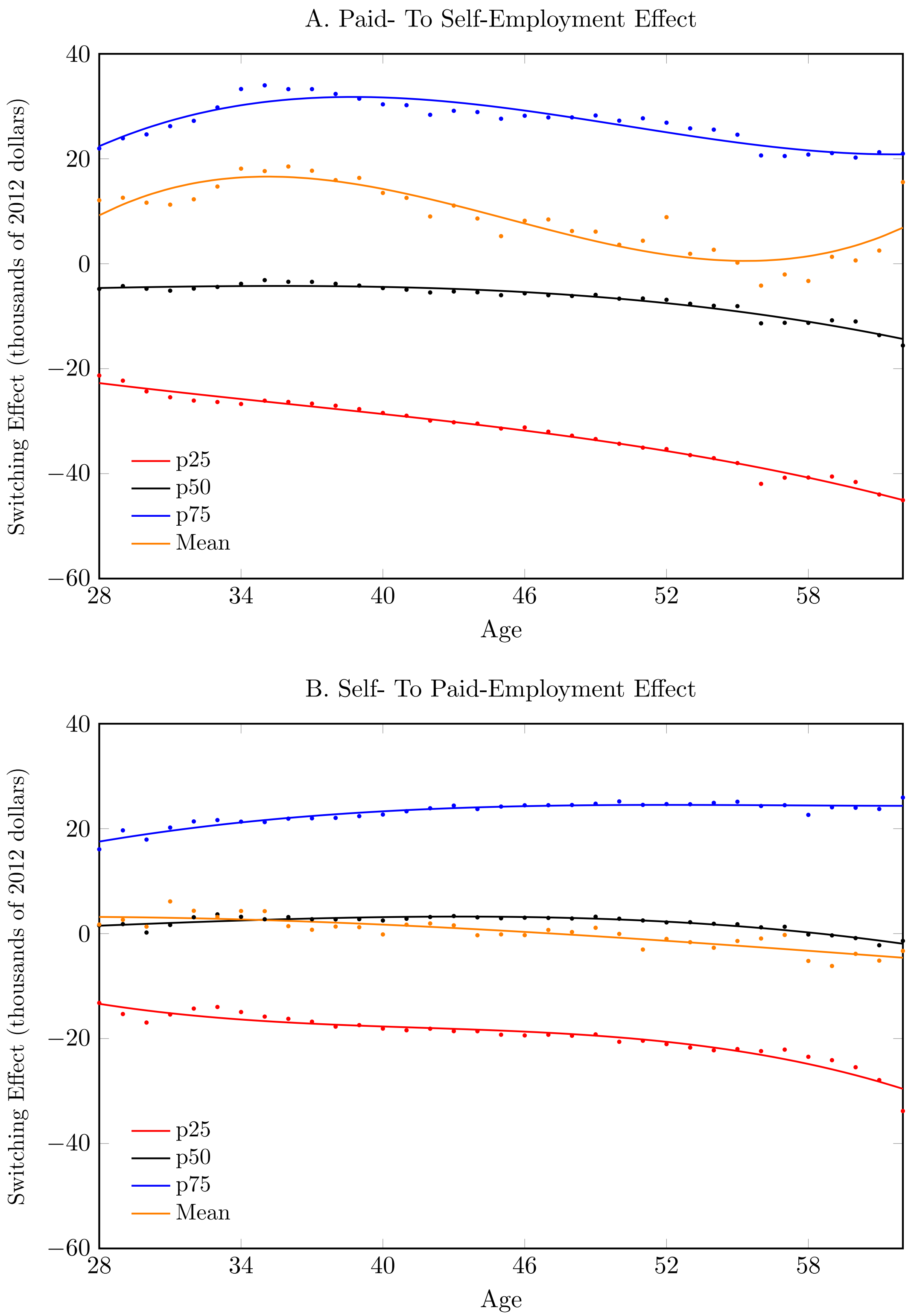
Figure 14: Difference in Past Income and Assets Among Switchers and Non-Switchers





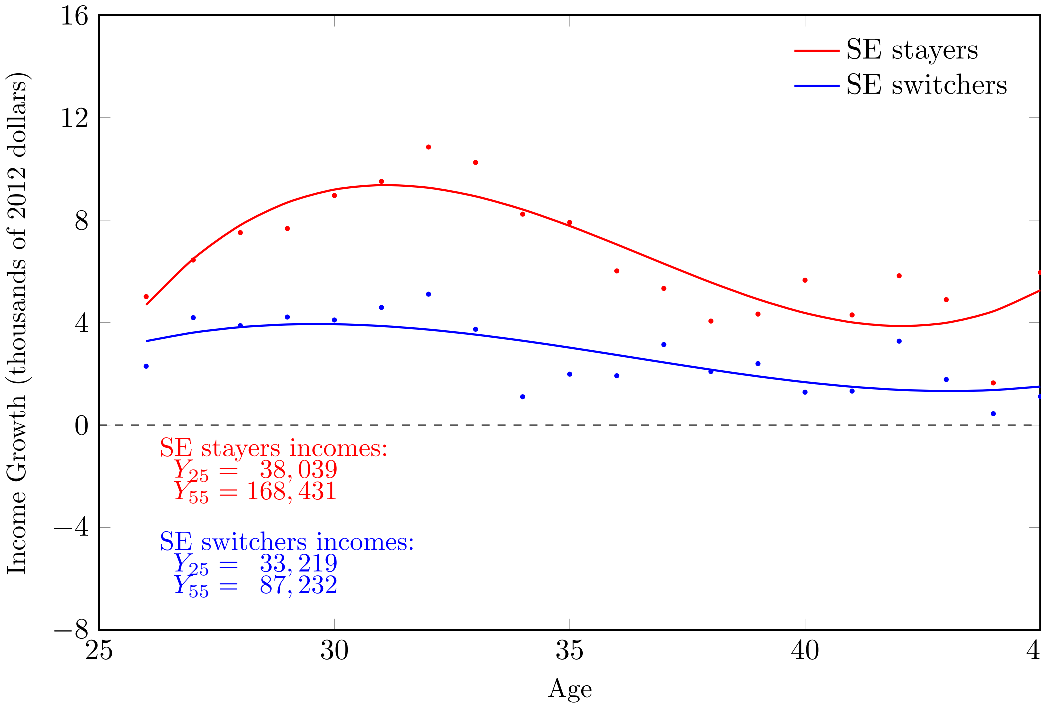
Note: The underlying sample is individuals with, at most, an observed single switch between paid- and self-employment. Panel A displays the interquartiles of the difference in past income by age for switchers and non-switchers and Panel B displays the differences in past asset income by age for switchers and non-switchers (see section 4 for details).

Figure 15: Switching Treatment Effects



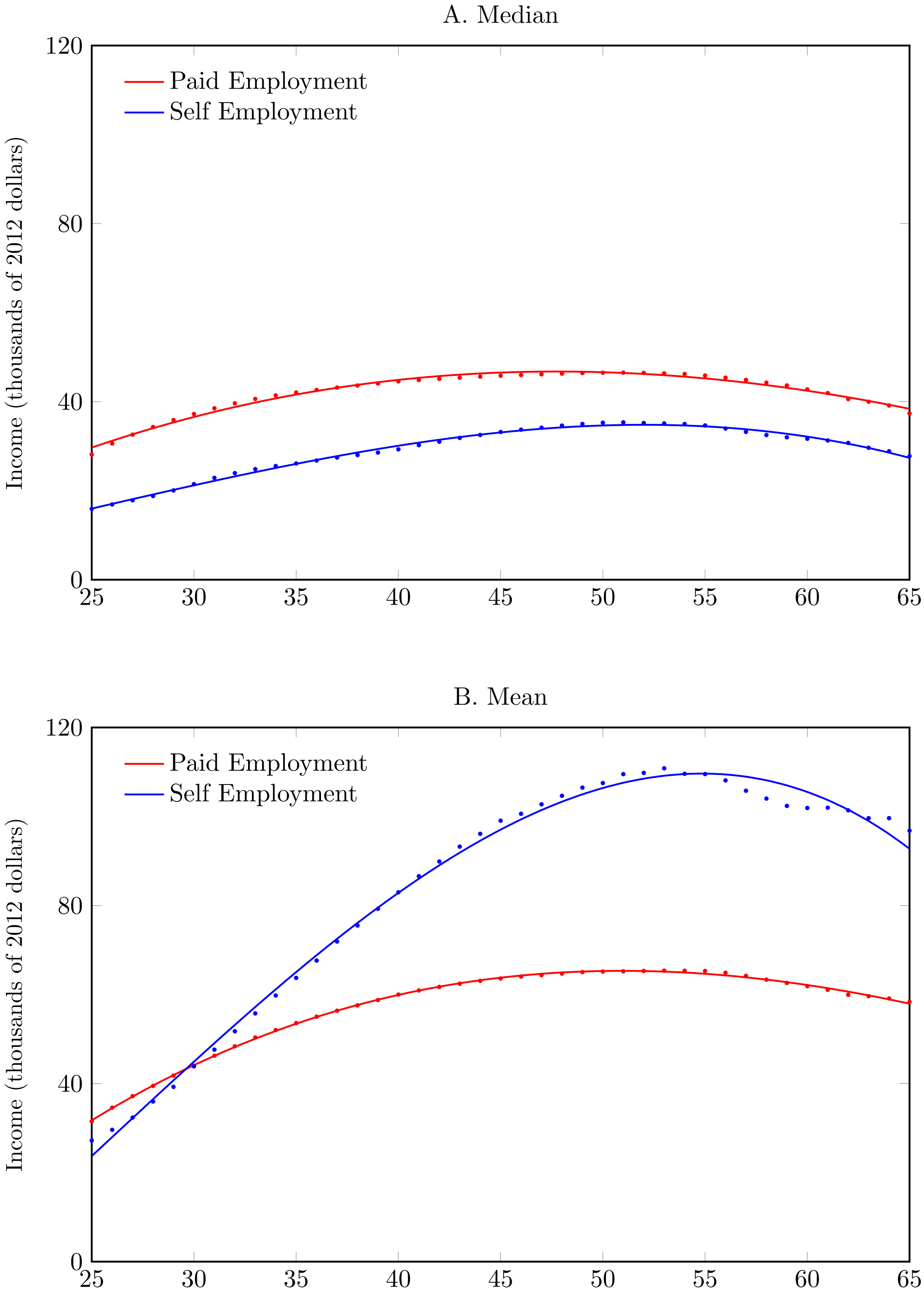
Note: The underlying sample is all individuals for whom we observe at least one switch between paid- and self-employment. Panel A shows the effect of switching from paid- to self-employment. Panel B shows the effect of switching from self- to paid-employment.

Figure 16: Growth Profiles of Young Entrepreneurs



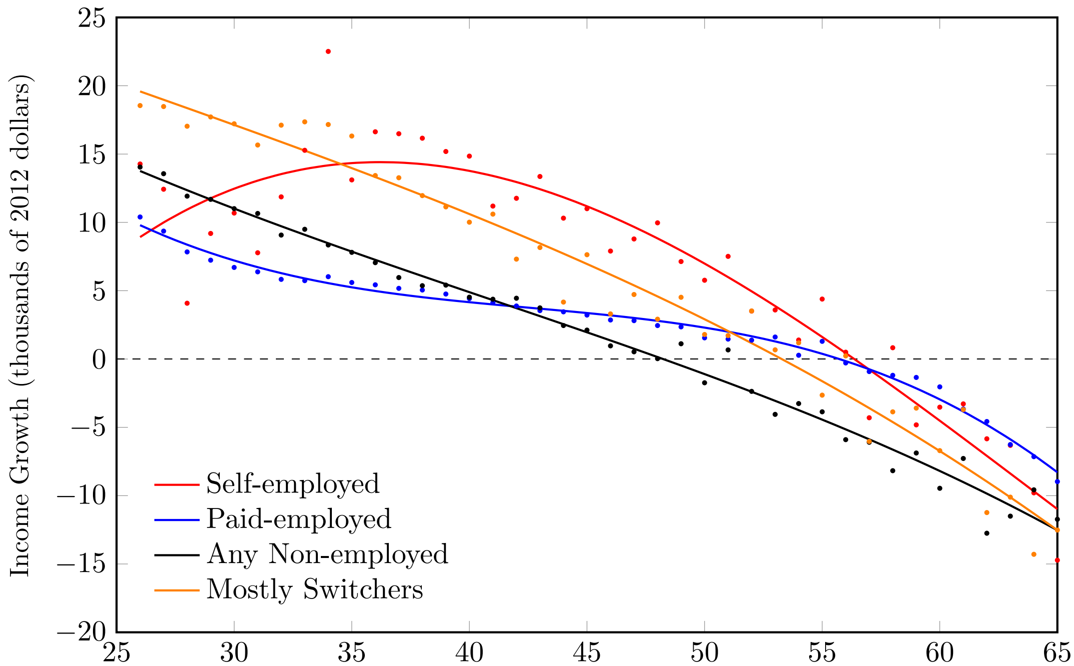
Note: The sample underlying these graphs are the attached self- and paid-employed groups. The figure shows the age effects, $\{\bar{\gamma}_g^a\}$, for young entrepreneurs (individuals with at least five years of self-employment experience before age 35) who stayed in self-employment or switched to paid-employment after age 35.

Figure 17: Emperical Moments By Employment Status



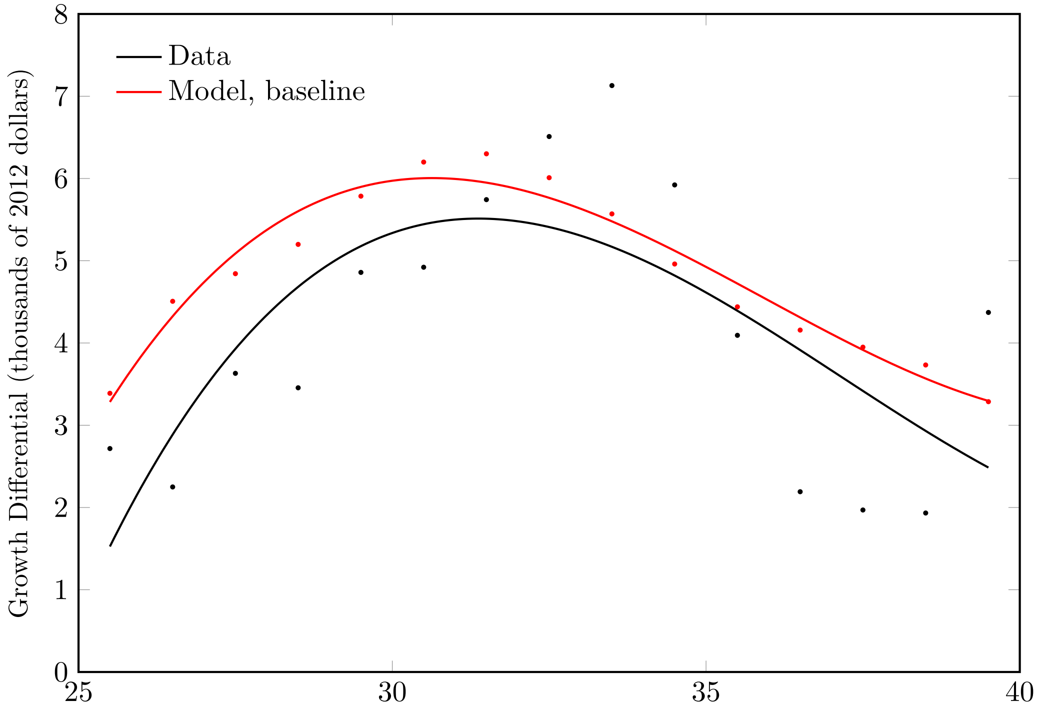
Note: The sample underlying these graphs are individuals who are classified as paid- or self-employed at a specific age. The figures show the median (Panel A) and mean (Panel B) of individuals' total income by age.

Figure 18: Top Income Bin Growth Rates



Note: The underlying sample is all individuals in the top 25% of average income. The figures show the age effects, $\{\bar{\gamma}_g^a\}$, for the invariant employment groups.

Figure 19. Growth Differentials for Young Entrepreneurs



Note: See Figure 16 for description of data.

Figure 20. Model Predictions as Intangible Revenue Share Varied