Where are the Young Entrepreneurs?
A Study of Entrepreneurship over the Life Cycle

Andrés Hincapié*
Department of Economics, Washington University in St. Louis

THIS VERSION: January 25, 2017

Abstract

Academics and the popular press celebrate entrepreneurship, particularly young entrepreneurs. However, most individuals do not start a business and, if they do, tend to do so well into their thirties. While policies encouraging young, would-be entrepreneurs are popular, little is known about whether they are effective. Using data from the Panel Study of Income Dynamics, I estimate a dynamic Roy model with imperfect information about ability to evaluate the relative importance of various economic determinants of entrepreneurial participation. Risk-averse, forward-looking individuals sequentially select entrepreneurial and paid-employment occupations based on their returns to experience, information value, non-pecuniary benefits, and entry costs. Results show that the main barriers faced by young entrepreneurs are entry costs and information frictions. I consider two policy counterfactuals: a subsidy targeting entry costs and entrepreneurship education targeting information frictions. I extend previous literature providing a mapping from the information quality of entrepreneurship education into career choices and long-term outcomes. A subsidy for young entrepreneurs increases participation but has small long-term effects. Entrepreneurship education can have sizable effects on participation and present value of income flows, even for low information quality. Nevertheless, the value of any particular entrepreneurship education program will depend on its cost and its information quality.

Keywords: Entrepreneurship, Occupational Choice, Correlated Learning.
JEL Classification: J23, J24

*E-mail: andres@wustl.edu. Address: Washington University in St. Louis, One Brookings Drive, St. Louis, MO, 63130. I thank my dissertation advisors Barton Hamilton (chair), George-Levi Gayle, Robert Miller, and Limor Golan for their helpful comments and invaluable advice. I have also received helpful comments from Brian Bergfeld, Chaoran Chen, Robert Munk, and Nicholas Papageorge. The paper have also benefited from presentations at Washington University in St. Louis (Department of Economics, Olin Business School, 2016 Econ Grad Student Conference), the Centre for European Economic Research (ZEW), and the 2016 EconCon at Princeton. I also thank the Center for Research in Economics and Strategy (CRES) at the Olin Business School for financial support during part of this research. All errors are mine.

1 Introduction

Entrepreneurship has long been considered an engine of innovation, growth, and development. These connections, already suggested by Schumpeter (1911) and more recently described by Baumol (2002), have found empirical support in the literature (Guiso et al. (2004), Chatterji et al. (2013)). In particular, young entrepreneurship is routinely celebrated by the popular press, as seen in the *30 under 30* collection by *Forbes* magazine, and by policy-makers, who highlight the economic potential gains from fostering young entrepreneurs OECD (2013). However, most people do not start a business during their careers and, if they do, tend to do so well into their thirties.[1] Although there are many policies attempting to stimulate entrepreneurship among individuals entering the labor market, it is not clear to what extent they induce young people to start businesses, what types of entrepreneurs they attract (e.g., high or low ability), and what are the long term consequences of these policies.[2]

In this paper, I develop a life-cycle framework analyzing entrepreneurial entry and exit based on the Roy (1951) model of occupational choice, and focusing on the self-employment decisions of young labor market entrants. In the model, risk-averse individuals sequentially choose salaried and entrepreneurial occupations based on their returns to experience, information value, uncertainty, non-pecuniary benefits, and entry costs. In particular, individuals decide to become entrepreneurs based on their beliefs about their own abilities. This paper adds to the literature by quantifying the relative importance of economic forces that determine whether and when people become entrepreneurs. Three elements are particularly novel to the empirical literature on entrepreneurship: first, an analysis of the timing of the choice; second, an assessment of the relative importance of risk aversion in a dynamic setting; and finally, an evaluation of the role of cross-occupation learning between salaried and entrepreneurial occupations.[3] In this framework, cross-occupation learning refers to the ability to transfer skills learned in one occupation to another (Liang et al. (2014)), as well as the ability to learn about one's entrepreneurial ability from paid-employment success.

A number of economic forces have been suggested in the literature to explain why individuals, both young and old, attempt entrepreneurship. Learning-by-doing, commonly characterized as experience accumulation that increases productivity, is one such force (Kawaguchi

[1] According to data from the Panel Study of Income Dynamics.

[2] Examples of such initiatives include the Thiel Fellowship (`http://thielfellowship.org`), the National Young Entrepreneurship Challenge (`https://www.nfte.com`), the Veale Young Entrepreneurship Forum (`http://vealeentrepreneurs.org`), and the Small Business Administration's Learning Center (`https://www.sba.gov/tools/sba-learning-center/training/young-entrepreneurs`).

[3] Munk (2015) approaches the timing of the entrepreneurial choice in a reduced-form setting revisiting the question of whether self-employment pays.

(2003), Lazear (2005), Lafontaine and Shaw (2016)).[4] Learning-by-doing predicts that individuals are more inclined to try entrepreneurship if they do not have to climb the productivity ladder first. Learning about one's entrepreneurial ability also determines entry.[5] If individuals are uncertain about their entrepreneurial ability, but the performance of their business helps them learn about it, they attempt entrepreneurship as long as their prior variance is high—they will want to learn whether they are in the "right" part of the distribution.[6] The option value of entrepreneurship, which intersects both learning-by-doing and learning about ability, also affects entry. Individuals attempt entrepreneurship because they can always switch back to paid employment if they discover that entrepreneurship is not the best option for them (Manso (2014) and Dillon and Stanton (2016)). However, reluctance to experiment increases if failure is penalized (Gottlieb et al. (2016)).

Risk aversion, credit constraints, and non-pecuniary benefits also affect the entry margin. Risk aversion pushes individuals away from entrepreneurship because it is a more uncertain occupation (Iyigun and Owen (1998)). Credit constraints in starting a business or in reaching optimal scale prevent less affluent individuals from trying their luck as entrepreneurs (Evans and Jovanovic (1989), Hurst and Lusardi (2004), and Buera (2009)). Finally, non-pecuniary motivations such as "being one's own boss" and "wanting flexibility over schedule" could also provide incentives for entry (Hamilton (2000), Hurst and Pugsley (2011, 2015)).[7]

More importantly, some of the forces explaining entry have predictions for whether entrepreneurs will be young. For instance, learning-by-doing implies that individuals who want to become highly productive entrepreneurs should start at an early age. Learning about ability gives the same prediction. High ability variance in entrepreneurship encourages individuals to seek to discover their place in the distribution as early as possible (Miller (1984)). Credit constraints preclude young individuals with weaker credit histories and less disposable wealth from entering. Risk aversion has a less clear effect on the timing of entry. Overall, risk aversion will prevent entry at any stage in an individual's career. However, if learning about ability reduces uncertainty over time (e.g. if learning is Bayesian), then the effect of risk aversion on entry can be attenuated as individuals acquire more experience. If success in paid employment correlates with higher entrepreneurial ability, favorable outcomes in paid employment may be associated with switching into self-employment. However, if paid-employment outcomes are uninformative of one's entrepreneurial ability, successful

[4] A well known example in the occupational choice literature is Keane and Wolpin (1997).

[5] Models of occupational choice with incomplete information include Jovanovic (1979), Sanders (2010), and Antonovics and Golan (2012).

[6] The high variation observed in entrepreneurial income suggests this could be the case.

[7] Other determinants explored in the literature include peer-effects (Nanda and Sørensen (2010)) and personality traits (Hamilton et al. (2016) and Humphries (2016)).

workers would be more willing to stay in paid employment not only because they seem better at it, but also because there is less uncertainty about the fact that they are.

There are several econometric challenges for estimation of the framework developed here. Given that individuals select occupations based on their beliefs, selection bias cannot be accounted for using first-differences estimators of occupation-specific income equations and panel data (Gibbons et al. (2005)).[8] Therefore, this paper accounts for selection bias using the likelihood function implied by the model. However, this econometric decision can come at the cost of multidimensional integration over unobserved ability vectors. To get around the computational burden, estimation of the parameters of the model follows a two-stage procedure using a combination of an Expectation-Maximization (EM) algorithm and a conditional choice probabilities (ccp) estimator (Hotz and Miller (1993), Arcidiacono and Miller (2011), James (2011)). The EM algorithm in the first stage bypasses the need for multidimensional integration. The ccp estimator in the second stage allows for a flexible treatment of the large state space of the problem, which includes continuous beliefs and experience for each occupation. The ccp estimator delivers such flexibility because the structural parameters can be estimated without solving the dynamic optimization problem at every candidate parameter vector in the search algorithm.

In the empirical analysis, this paper uses data from the Panel Study of Income Dynamics (PSID). The sample is restricted to white and black men between the years 1968 and 1996. Moreover, following suggestive evidence in Levine and Rubinstein (Forthcoming), entrepreneurship is disaggregated by incorporation status. Entrepreneurs with promising abilities benefit more from the incorporated organizational form because it encourages entrepreneurial risk-taking by offering limited liability and by facilitating fund-raising through the issuance of stock. Less promising entrepreneurs, on the other hand, benefit more from the less complex unincorporated form that offers lower administrative costs and regulatory burden. Results using the PSID suggest that incorporated entrepreneurs are more similar to white collar workers than they are to unincorporated entrepreneurs.

The framework is used to quantify the importance of the economic forces at play by comparing the baseline model against counterfactual regimes that shut down each economic force. Results indicate that learning-by-doing and entry costs have the largest effects preventing individuals from attempting entrepreneurship. The role of learning-by-doing is evaluated by turning the profile of returns to experience into a flat average return. Given that the profile of returns to experience in entrepreneurship is steep, the flat average in the counterfactual is

[8]Since individual's beliefs change over time as they acquire more information, their ability "does not translate into a fixed effect" in an income equation (Gibbons et al. (2005)). Cross-section data are even more problematic as they provide no historic information to model the process of belief formation.

high, making individuals more willing to experiment. Risk aversion and information frictions also play important roles. For instance, shutting down risk aversion increases the percentage of individuals who attempt incorporated entrepreneurship by 40% and eliminating information frictions increases this number by 35%. Eliminating cross-occupation learning reduces the proportion of individuals attempting entrepreneurship by 10%. Although these effects seem small, in the long term they are stronger: eliminating cross-occupation learning decreases the present value of income (PVI) of incorporated entrepreneurs by about a quarter.

Results also show that the two main barriers to young entrepreneurship are entry costs and information frictions. In the model, entry costs capture barriers to entrepreneurship not explicitly modeled, such as credit constraints.[9] In order to make the link between entry cost and credit constraints, these costs are interacted with age and a permanent wealth component (estimated as a fixed effect outside of the model). Estimates show that younger individuals as well as individuals with lower permanent wealth face higher barriers to entry. Flattening the profile of entry costs, allowing individuals of all ages to face the same average entry cost, closes the gap in average first-entry age between white collar work and entrepreneurship by about 70%. Eliminating information frictions, providing full information about ability, induces individuals to enter entrepreneurship earlier closing the first-entry age gap by 20%.

Focusing on the main barriers to young entrepreneurship, the paper undertakes two policy counterfactuals: a subsidy that targets entry costs and entrepreneurship education that targets information frictions.[10] Previous literature has shown that entrepreneurship education can shift beliefs (von Graevenitz et al. (2010), Oosterbeek et al. (2010)). The paper takes this result as given and extends the literature providing a mapping from movements in beliefs, generated by entrepreneurial education of a given quality, into career choices and long-term outcomes. Results suggest that a blanket subsidy for young incorporated entrepreneurship increases participation and has a small positive effect on the average PVI net of the subsidy.[11] Additionally, incorporated entrepreneurship education can generate sizable increases in young incorporated entrepreneurship and PVI even at low levels of information quality. Nevertheless, caution must be taken when reading these results. The information quality of any specific policy may be different and its cost may well exceed the additional income it generates.

The rest of the paper is organized as follows. Section 2 presents the data and describes the main regularities motivating the research question and modeling choices. Section 3 describes

[9]Unfortunately, the PSID lacks wealth data for most years during the period studied.

[10]Young entrepreneurs are defined as those who attempt entrepreneurship during the first five years of their labor market careers.

[11]Assuming that the incorporated self-employed are closer than their unincorporated peers to what is commonly thought of as "the entrepreneur," these counterfactuals focus on incorporated entrepreneurship.

the characteristics of the model and its implications. Section 4 describes the estimation method. Section 5 discusses the estimated parameters and presents the decomposition exercise. Section 6 introduces the policy counterfactuals. Section 7 concludes.

2 Data

The Panel Study of Income Dynamics (PSID) is used in the empirical application.[12] There are three main reasons that motivate this choice. First, many individuals in the PSID are observed at the onset of their labor market careers and yearly from that point forward. This allows for the construction of measures of accumulated occupation-specific experience and occupational income at any point in a respondent's career. Second, whenever an individual declares himself to be self-employed, the survey's questions on self-employment allow for disaggregation into incorporated and unincorporated self-employment. Multiple differences between these self-employment alternatives have been previously identified by Levine and Rubinstein (Forthcoming). Finally, although not available for most periods, the PSID collects data on wealth that will prove useful in the analysis of entry costs.

The sample is restricted to white and black men between years the 1968 and 1996. It contains survey information on occupation, self-employment status, business ownership, incorporation status, labor income, business income, working hours, completed education, age, race, marital status, and wealth.[13] Individuals' labor market careers are set to start at the beginning of their potential experience.[14] Both types of occupations, salaried and entrepreneurial, are disaggregated to exploit differences in returns to experience as well as differences in terms of the information they provide. For paid employment, the 3-digit occupation code is used to generate two categories: blue collar and white collar.[15] This aggregation has been used previously in occupational choice models studying paid employment (Keane and Wolpin (1997)). For self-employment, which is interchangeably refered to as entrepreneurship in this study, the incorporation status questions in the survey are used to create two categories: unincorporated and incorporated.

[12]The study started in 1968 with a representative sample of about 18,000 individuals in 5,000 families in the United States. Information about these individuals and their descendants was collected yearly until 1996, after which the study became biennial.

[13]Biennial data collected after 1996 was not used, since it would require making assumptions about occupational choices and income in years where no data were collected.

[14]Potential experience starts at the end of their completed education. All individuals are assumed to start school at age 6. The Data Appendix contains a detailed description of how each variable is constructed.

[15]The PSID provides 3-digit occupation codes from the 1970 Census of Population, constructed using the Alphabetical Index of Industries and Occupations issued June 1971 by the U.S. Department of Commerce and the U.S. Census Bureau.

The income measure for paid employment corresponds to the individual's reported annual labor income. For self-employment, measuring income is less transparent. Since incorporated individuals are not asked about their business income in the survey, their reported labor income is used as their income measure.[16] For unincorporated individuals, who are not sheltered from the losses of their ventures through limited liability, the measure corresponds to the sum of the quantity reported as labor income plus the quantity reported as business income. Income measures are converted to hourly rates by dividing annual income figures by reported annual working hours.

The PSID includes a measure of wealth for selected years starting in 1984. It is constructed as the sum of six types of assets (farm business, checking or savings accounts, real estate other than main home, stocks, vehicles, and other assets) net of debt value plus value of home equity. The survey does not include data on wealth at every period. Therefore, in the current analysis a measure of permanent wealth will be considered instead of the separate wealth observations. The measure of permanent wealth, denoted ω_i, is obtained as the constant plus the fixed effect of a regression of wealth on a second degree polynomial on age. In estimation, only individuals with at least three wealth data points are considered. The variable ω_i is meant to capture long-run differences in access to resources.

After dropping observations of individuals who lack data on relevant variables, the analytic sample contains 1,506 individuals and 21,334 individual-year observations. Table 1 shows that about one-fifth of the sample is African American and 42% of the individuals have college education or more.[17] The average entry age to the labor market is about 22 years. Finally, the average permanent wealth, is \$400,000 (measured in year 2000 USD) with a large standard deviation of \$674,000.[18] In addition, consistent with the higher complexity of the incorporated organizational structure, Table 2 shows that incorporated individuals tend to be more educated than their unincorporated counterparts. Finally, self-employed individuals, especially incorporated entrepreneurs, are more likely to be married than paid employees.

[16] This measure corresponds to what Hamilton (2000) terms "the draw" or the difference between net profit and retained earnings. The characteristics of corporations as separate legal entities from the business owners justify the use of this measure.

[17] Notably, for the period of study, the proportion of individuals with college education in the sample is higher than what it was for the U.S. adult population, 22% (Ryan and Siebens (2012)). The disparity arises when the selection criteria imposes that individuals must be observed from the beginning of their labor market careers and widens more once observations without enough wealth data points are dropped. The lower the education the earlier they would enter the job market and the less likely the PSID is to observe them from the beginning of their careers.

[18] As mentioned above, the measure of individual permanent wealth corresponds to the constant plus the individual fixed effect of a regression of wealth on an age profile. Hence, the level of this measure depends on the shape of the polynomial implemented. The estimated value of the constant is about \$418,000 (see Data Appendix).

2.1 Incorporated and Unincorporated Entrepreneurs

Using the individual's incorporated status, this paper distinguishes between two types of self-employment. This disaggregation follows Levine and Rubinstein (Forthcoming) who introduce the differences between the organizational forms of unincorporated and incorporated businesses and the differences between the individuals they attract. They highlight how the organizational form of corporations facilitates growth and risk-taking behavior. Individuals seeking to establish businesses with high potential for development tend to be more attracted to this organizational form.

There are three distinctive characteristics of incorporated businesses. First, they are separate legal entities from their owners. This allows the corporation to own property, carry on business after the death of its owners, incur liabilities, and sue or be sued. Importantly, it also means that corporations can operate isolated from sudden shocks in an owner's personal finances. Second, corporations have limited liability against creditors. In other words, creditors seeking debt repayment can go after a shareholder's assets only to the extent of her investment in the business. This is precisely one of the reasons that motivates investors and venture capitalists to invest. Instead, unincorporated businesses owners have their personal assets exposed to the losses of their business. Third, corporations can issue shares of stock. This makes it easier for them to raise money in order to develop the business. It also makes transferability of ownership simpler than for sole proprietors or general partners.

But the advantages of incorporating a business come at the costs of more complex administrative activities, higher administrative costs, and potentially higher taxes.[19] Consequently, non-employer self-employed individuals and other small business owners will find incorporation unattractive (e.g. individual construction contractors, car repair shop owners). They are less likely to develop the business much further, so they have less incentive to incorporate. In contrast, incorporated self-employed individuals seek to take advantage of the organizational structure of the corporation to grow and develop their businesses.

The organizational forms of these two types of businesses also suggest that the abilities required from individuals trying to sort into them and the skills developed while working in each of them may be different. This observation, which further motivates the separation made in this paper, finds empirical support in the data. Table 3 shows that entrepreneurs moving into paid employment differ in terms of the occupations they switch to. Unincorporated entrepreneurs are just as likely to switch to either blue collar or white collar work,

[19]These activities include holding annual meetings, recording meeting minutes, and keeping up at all times corporate documents such as the register of directors, the share register and the transfer register. Additionally, corporations are taxed and their owners are also taxes on the dividends.

while incorporated entrepreneurs tend to transition to white collar work. Further differences between these two types of self-employment will be introduced in the preliminary analysis of the data below. Notably, it will be shown that successful white-collar workers, characterized by higher residual income, are more likely to switch into incorporated entrepreneurial activities.[20] In general, the data suggest that incorporated entrepreneurs look more like white collar workers than unincorporated entrepreneurs.[21]

2.2 Preliminary Analysis of the Data

Several stylized facts in the data motivate this research and the dynamic model of occupational choice with learning used for analysis. Most individuals do not attempt entrepreneurship during their careers and they are even less likely to start their careers as entrepreneurs—individuals entering entrepreneurship tend to be older and have accumulated some paid employment experience prior to entry. Besides, those who attempt entrepreneurship tend to transition out of it faster than those who enter paid employment. In terms of hourly income, entrepreneurial occupations display higher variation than paid employment occupations, even after controlling for observables. Finally, successful white collar workers, characterized by higher residual income, are more likely to switch into incorporated entrepreneurial activities than their less successful peers.

Entrepreneurship is less common in an individual's career than paid employment. Table 4 shows that the proportion of individuals who attempt entrepreneurial occupations is less than half the proportion of individuals who attempt salaried occupations.[22] Furthermore, most of the 4,294 occupational spells in the sample occur in paid employment occupations and they are more than 60 percent longer than spells in entrepreneurial occupations (see Table 5). This is consistent with the transition patterns in Table 3, as salaried occupations tend to be more absorbing than entrepreneurial occupations.

Table 4 introduces the puzzle of young entrepreneurship. It shows that those who attempt entrepreneurial occupations tend to do so later in their careers, after accumulating more than 8 years of paid-employment experience on average. This translates into a gap of

[20]Residual income in occupation-specific regressions after controlling for demographics and a quadratic in occupation-specific experience.

[21]In the PSID, self-employed individuals, as well as paid employees, report occupation and industry. In computations not shown here, incorporated individuals are about 20 percent points more likely to report working in white collar occupations than unincorporated individuals. On the other hand, unincorporated entrepreneurs are about 20 percent points more likely to report belonging to the construction and repair industry than incorporated entrepreneurs.

[22]In separate calculations, the percentage of individuals who try at least one type of entrepreneurship by age 50 is about 34%. Virtually everybody in the sample tries paid employment

more than 7 years in average entry age between entrepreneurial occupations and salaried occupations. This gap runs opposite to the prediction of a parsimonious model of uncorrelated learning about ability (Miller (1984)) and does not seem suggestive of individuals having to climb the productivity latter early on to become highly productive entrepreneurs. Notably, although few individuals start their careers as entrepreneurs (see Table 5), Figure 1 shows that participation increases as individuals age.

In terms of hourly income, entrepreneurial occupations display higher variation across observations than paid employment occupations. This is shown in Table 2, where individual-year observations are summarized by occupation. In particular, the variance of hourly income in incorporated entrepreneurship is more than three times as large as the variance in white collar work. This difference remains even after controlling for demographics and occupation-specific experience. Interestingly, even though unincorporated entrepreneurship has a higher hourly income variance than white collar work, they share similar average hourly income. Incorporated entrepreneurship, however, has a mean hourly income that is 75% higher.

Finally, successful white-collar workers, characterized by higher residual income, are more likely to transition into incorporated entrepreneurship than their less successful counterparts. This is shown in Figure 2. On the x-axis is the quintile of average residual income at t. On the y-axis is the probability of switching into the two types of entrepreneurship at $t+1$. Higher residual income in either salaried occupation is generally associated with a smaller probability of switching into unincorporated entrepreneurship. This is consistent with uncorrelated learning, where unexplained success is only informative of ability in the current occupation. However, successful white-collar workers, as measured by higher residual income, are more likely to switch into incorporated entrepreneurship than their less successful peers. This is consistent with correlated learning about ability between white collar work and incorporated entrepreneurship, and is also consistent with the similarities between incorporated entrepreneurs and white collar workers mentioned in the previous section.

These facts are interpreted through the lens of a dynamic model of occupational choice with accumulation of experience (learning-by-doing) and learning about ability. Learning about ability is often modeled using Bayesian decision makers that draw information from their labor market outcomes and combine it with potentially heterogeneous priors in order to update their beliefs about their aptitude at various firms or occupations (Jovanovic (1979), Miller (1984)). A prediction of these models, that the separation probability of a worker is decreasing in her tenure on the job, is found in Figure 3 at the occupation level.[23] Notably, the probability of switching from entrepreneurship decreases at a much higher rate than it

[23]This relation partially results from workers being less uncertain of their competence, which ameliorates the effect of new information.

does for salaried occupations during the first five years of accumulated occupation-specific experience.

The stylized facts also motivate the introduction of cross-occupation learning. Cross-occupation learning-by-doing aims to capture the nature of the mix of salaried experience that individuals acquire before entering each entrepreneurial occupation (see Table 4). Moreover, cross-occupation learning about ability (correlated learning) aims to capture the relationship between unexplained success in white collar work and entry into incorporated entrepreneurship.

3 Model

In the model, forward-looking, risk-averse individuals face dynamic incentives that reflect two processes: accumulation of experience (learning-by-doing) and accumulation of information (learning about unobserved ability). The model captures the transferability of acquired skills as well as spillovers of information. For instance, a financial manager who decides to become an entrepreneur later in his career may transfer the managerial skills he has acquired into his business. Additionally, his success or failure as a manager may reveal his entrepreneurial ability as well.

Individuals enter the labor market immediately after finishing their education. Their productivity in each occupation is determined by their experience, their unobserved ability, and idiosyncratic shocks that prevent them from learning their ability immediately after the first period. Therefore, they receive noisy income signals that they use to update beliefs about their own ability. Individuals maximize their expected utility using their updated beliefs to compute expectations. They have differential preferences for occupations and are able to smooth consumption over time.

3.1 Occupations and Individual Characteristics

Immediately after finishing his education, individual i of age t_{i0} enters the labor market.[24] He then decides which occupation to join and how much to consume. If he decides to work, he can be a white collar or blue collar paid employee, or he can be an unincorporated or incorporated self-employed. Denote $d_{kit} \in \{0, 1\}$ as an indicator of whether or not he chooses alternative k at age t.

[24]The empirical analysis uses data for males only. Therefore, masculine pronouns are used to describe the model.

Every individual has a vector of observable characteristics, h_{it}, that partly determines his productivity in each occupation. h_{it} includes his race (white or black), his education level (high school or less, some college, college, and more than college), as well as observable characteristics that change over time (age, marital status, and accumulated experience). He has perfect foresight over his marital status. His experience is collected in a 4-dimensional vector, x_{it}, that contains his accumulated experience in every occupation. He starts his career with no experience in any occupation and from that point forward the kth component of his experience vector evolves as a function of his choices as follows

$$x_{kit+1} = x_{kit} + d_{kit} \tag{1}$$

3.2 Preferences

The individual is infinitely-lived and discounts next period's utility by the factor β. He works until age T and begins his retirement at age $T + 1$.[25] After reaching retirement age, he only decides how to smooth his remaining savings. In order to capture the effects of more uncertain entrepreneurial outcomes on occupational choices, while retaining tractability, the individual's flow utility is characterized by a CARA function of consumption, c_{it}, with absolute risk aversion parameter ρ. His lifetime utility at period $t \leq T$ is given by

$$-\sum_{s=t}^{T}\sum_{k=0}^{4} \beta^{s-t} d_{kit}\alpha_{kit}(h_{it}) \exp\{-\rho c_{it} - \varepsilon_{kit}\} - \sum_{s=T+1}^{\infty} \beta^{s-t} \exp\{-\rho c_{it}\} \tag{2}$$

The marginal contribution of consumption to his utility is occupation-specific and is determined by the non-pecuniary cost of each occupation, α_{kit}, which is allowed to vary by education level. Given an education level, the non-pecuniary cost is a function of his vector of observables

$$\alpha_{kit}(h_{it}) = \exp\{\alpha_{k0} + \alpha_{k1} black_i + \alpha_{k2} married_{it} + 1\{x_{kit} = 0\}(\alpha_{k3} + \alpha_{k4} t + \alpha_{k5}\omega_i + \alpha_{k6} t\omega_i)\} \tag{3}$$

Notably, α_{kit} includes a first-time entry cost which is a function of permanent wealth (ω_i) and age. This is a reduced-form way of capturing barriers to entry that are not explicitly modeled. For entrepreneurship, the age profile of the entry costs is meant to capture the difficulties that young individuals with weaker credit histories and less savings may face. Additionally, the introduction of the permanent wealth measure is meant to capture whether more affluent individuals, in a long-term sense, face smaller entry costs to entrepreneurship.

[25] Retirement age is set at $T + 1 = 51$ for data availability reasons.

For identification reasons, the non-pecuniary cost of not working, α_{0it}, is normalized to one. Finally, ε_{kit} is the kth component of a 5-dimensional vector of choice-specific taste shocks that he observes before choosing an alternative. The taste shocks are unobserved to the econometrician and are assumed to be drawn from a Type-I Extreme Value distribution, independent across individuals, periods, and alternatives.

3.3 Income and Learning

Individual i of education level s starts his labor market career endowed with a vector of occupation-specific abilities $\mathcal{M}_i = \{\mu_{1i}, \ldots, \mu_{4i}\}$ drawn from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Delta}_s$.[26] His ability partially determines his productivity in each occupation. His hourly income at the beginning of period $t+1$, from choosing occupation k at period t, is given by

$$y_{kit+1} = f_k(h_{it}; \theta_k) + \mu_{ki} + \eta_{kit+1} \tag{4}$$

His hourly income is the sum of an idiosyncratic productivity shock η_{kit+1}, his unobserved occupation-specific ability μ_{ki}, and a function of his observable characteristics $f_k(\cdot)$, which is characterized by a vector of parameters θ_k.[27] Productivity shocks η_{kit+1} are distributed $N(0, \sigma^2_{\eta_k})$ and are independent over time and across individuals and occupations.

The individual does not observe μ_{ki} and η_{kit+1} separately, which prevents him from learning his ability immediately. Instead, he observes their sum, $\zeta_{kit+1} = y_{kit+1} - f_k(h_{it}; \theta_k)$ after choosing occupation k.[28] He follows Bayes' Rule and uses the information he has acquired, i.e. his residual income signal ζ_{kit+1}, to form beliefs $\mathbb{B}_{it+1}$ about his ability. At the beginning of his career, he believes he is no different from any of his peers, and his prior beliefs correspond to the population distribution of ability for people with his education level. Therefore, by virtue of the joint normality of the distribution of ability, his initial beliefs can be characterized by the mean and variance of the population distribution: $\mathbb{B}_{it_{i0}} = \langle \mathbf{0}, \mathbf{\Delta}_s \rangle$. The normality of the prior and the idiosyncratic shocks yields a posterior which is also multivariate normal. Therefore, his beliefs at any period follow a normal distribution and can be characterized by a mean vector $\mathbb{E}_{it}$ and a covariance matrix $\mathbb{V}_{it}$.

Updating rules for similar problems have been previously obtained in the literature (De-

[26]The distribution is set around zero because it is not possible to identify the education level effect in the income equation as well as the mean of the distribution, as it depends on education.

[27]In particular, $f_k(\cdot)$ captures the returns to experience using step-functions of the experience vector x_{it}.

[28]In this sense, he is paid his actual productivity as opposed to his expected productivity. This assumption, although less compelling for paid employment, is a natural assumption for entrepreneurial income, which is not contracted upon.

Groot (1970), James (2011)). Define the 4-dimensional vector ζ_{it} with characteristic component $\zeta_{\{k\}it}$ and the 4×4 diagonal matrix Σ_{it} with characteristic component $\Sigma_{\{k,k\}it}$ as follows

$$\zeta_{\{k\}it} = \begin{cases} \zeta_{kit} \text{ if } d_{kit-1} = 1 \\ 0 \text{ otherwise} \end{cases} \qquad \Sigma_{\{k,k\}it} = \begin{cases} 1/\sigma^2_{\eta_k} \text{ if } d_{kit-1} = 1 \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

After receiving his income signal from last period's work, he updates his beliefs to

$$\mathbb{E}_{it} = \left[\mathbb{V}^{-1}_{it-1} + \Sigma_{it}\right]^{-1} \left[\mathbb{V}^{-1}_{it-1}\mathbb{E}_{it-1} + \Sigma_{it}\zeta_{it}\right] \tag{6}$$

$$\mathbb{V}_{it} = \left[\mathbb{V}^{-1}_{it-1} + \Sigma_{it}\right]^{-1} \tag{7}$$

These updating rules reflect how his beliefs change as a function of his experience and information received. Equation (6) implies that the effect of a very noisy signal on the prior mean is minor.[29] Moreover, equation (6) determines the extent to which learning about ability can happen across occupations. For instance, the direction and magnitude of the adjustment in beliefs of a white collar worker regarding his entrepreneurial ability is determined by the off-diagonal terms in the variance matrix $\mathbb{V}_{it}$. The larger these covariances are, the larger the adjustment will be.[30] Equation (7) implies that the prior variance at t is a deterministic map of the vector of accumulated experience x_{it} and the covariance matrix of the ability distribution (see Model Appendix). This implies that, conditional on x_{it}, the order in which the individual samples occupations prior to t is irrelevant to determining the posterior variance. More importantly, provided experience is already included in his state, it implies that $\mathbb{V}_{it}$ is redundant information.

3.4 Optimal Choices

At the beginning of any period before retirement, the individual receives his income from last period's occupation and observes his vector of taste shifters ε_{it}. Using his income signal, he updates his beliefs. Given that he can smooth consumption over time, he simultaneously chooses his consumption and asset portfolio, as well as whether to work and which occupation to join.[31] The set up for the consumption smoothing problem follows Margiotta and Miller (2000) and Gayle et al. (2015) whose main result will hold here: the indirect utility from optimal consumption has closed form, and the occupational choice will be independent of

[29] A noisy signal is characterized by high idiosyncratic variance $\sigma^2_{\eta_k}$.

[30] The marginal effect of a signal from occupation k on the next period's prior mean of occupation k' equals $(1/\eta_k)\mathbb{V}_{\{k,k'\}it}$.

[31] Upon retirement, he simply decides on his consumption and assets portfolio in order to smooth his remaining wealth.

disposable wealth. This result will substantially facilitate estimation and its adoption is also data-driven given the lack of wealth data in the PSID for most years in the period studied. Therefore, credit constraints are not explicitly modeled, and occupational choices will depend on relative differences in non-pecuniary benefits, expected flow payoffs given experience and beliefs, and continuation values from future experience and beliefs induced by each alternative.

3.4.1 Consumption

The set up of the consumption smoothing problem aims to relax the hand-to-mouth assumption that would force individuals to absorb the entire variation in income every period. The individual has access to a contingent-claims market for consumption goods to smooth his consumption using his wealth. However, income is assumed uninsurable to capture "unobservable insurance risk or unobserved levels of effort in labor supply" (Green (1987)).

Let λ_τ denote the derivative of the price measure for claims to consumption at date τ.[32] If he decides to work in occupation k, individual i supplies $\bar{L}_k$ hours inelastically.[33] Dropping the occupation indicator, working at age t yields him annual income $\bar{L}y_{it+1}$ at the beginning of $t+1$. Hence, the law of motion for his disposable wealth is:

$$E_t[\lambda_{t+1}\xi_{it+1}|d_{it}, h_{it}, \mathbb{E}_{it}] + \lambda_t c_{it} \leq \lambda_t \xi_{it} + E_t[\lambda_{t+1}\bar{L}y_{it+1}|d_{it}, h_{it}, \mathbb{E}_{it}]$$

where the expectation is conditional on his choice at period t, d_{it}. His budget constraint reflects his financial resources, which are allocated to current period consumption and next period savings.

Similar to results in Margiotta and Miller (2000), the assumptions on the market for consumption claims and the CARA nature of the flow utility yield an expression for the value function that can be separated into two factors: an indirect utility function for wealth and an index that captures the value of human capital and information. This expression, presented below, satisfies a portfolio separation property (Altuğ and Labadie (1994)): consumers will hold only a few securities. The first one is a bond b_τ that, contingent on the history through calendar date τ, pays a unit of consumption from period τ in perpetuity in date-τ prices. The second one is a security a_τ that pays the random quantity $(\ln \lambda_s - s \ln \beta)$ of consumption

[32] The commodity space for consumption goods is formed by consumption units at date 0 and claims to consumption at calendar date τ contingent on how history unfolds. λ_τ denotes the derivative of the price measure for claims to consumption at date τ, Λ_τ. Therefore, the price of a unit of consumption to be delivered with certainty at date τ in terms of date 0 consumption units is $E[\lambda_\tau]$.

[33] $\bar{L}_k$ is specified as as the average amount of hours worked by individuals in occupation k in the sample.

units from period τ in perpetuity, in date-τ prices. The prices of these assets are given by

$$b_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau}\right] \qquad a_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau}(\ln \lambda_s - s \ln \beta)\right] \tag{8}$$

Individuals in the model can accurately forecast the price of both assets. The state space of the individual's dynamic problem is then formed by his vector of observable characteristics, his beliefs, his wealth, and the prices of these assets. Let $\tau(t)$ be the calendar date when the individual is of age t and let ξ_{it} denote his disposable wealth at t. Following Margiotta and Miller (2000), and dropping the index i for simplicity, the value function solving his savings problem at retirement age $T+1$, in present value terms is

$$V_{T+1}(h_{T+1}, \mathbb{E}_{T+1}, \xi_{T+1}, a_{\tau(T+1)}, b_{\tau(T+1)}) = -\lambda_{\tau(T+1)} b_{\tau(T+1)} \exp\left(\frac{-(\rho\xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right) \tag{9}$$

His occupational ability as well as his experience become irrelevant once he retires.[34] In fact, as he receives no retirement flow income, his present value only depends on his remaining wealth and the price of the assets $a_{\tau(T+1)}$ and $b_{\tau(T+1)}$. At any age before retirement, he chooses an occupation in addition to his consumption and asset portfolio. Proposition 1 provides the ex-ante value function of an individual at any age before retirement.[35]

Proposition 1. *At any age t before retirement, $t \leq T$, the value function of an individual who has not yet observed his taste shocks, ε_t, can be written as*

$$V_t(h_t, \mathbb{E}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}) = -\lambda_{\tau(t)} b_{\tau(t)} \exp\left(\frac{-(\rho\xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right) A_t(h_t, \mathbb{E}_t) \tag{10}$$

where $A_t(h_t, \mathbb{E}_t)$ is defined recursively as

$$A_t(h_t, \mathbb{E}_t) = \\ \sum_{k=0}^{4} p_{kt}(h_t, \mathbb{E}_t) \alpha_{kt}(h_t)^{1/b_{\tau(t)}} E_\varepsilon[e^{-\varepsilon^*_{kt}/b_{\tau(t)}}] E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1} | \mathbb{E}_t, h_t]^{1-1/b_{\tau(t)}} \tag{11}$$

with $A_{T+1}(h_{T+1}, \mathbb{E}_{T+1}) \equiv 1$ and $v_{kt+1} \equiv \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right)$.

[34] More realistic models would have ability as well as accumulated human capital generating an income stream after retirement. This paper abstracts from such considerations, but acknowledging that retirement considerations would play a stronger role in such models as various occupational paths summarized by their experience vector would generate alternative retirement income flows.

[35] The ex-ante value function is defined as the value function before knowing the realization of the vector of taste shocks, ε_{it}

The probability of choosing k at age t conditional on characteristics and beliefs is denoted $p_{kt}(h_t, \mathbb{E}_t)$. ε^*_{kt} is the value of the taste shock ε_{kt} conditional alternative k being chosen at t. The deterministic transition from h_t into h_{t+1} is denoted $\bar{H}_{kt+1}(h_t)$, and the stochastic transition from yesterday's beliefs into today's is denoted $\mathbb{E}_{kt+1}$; both are conditional on choosing k at t.

Proof: See Model Appendix.

The ex-ante value function in (10) can be separated into two factors: the indirect utility from wealth and an index that represents the value of accumulated experience and information. The index $A_t(h_t, \mathbb{E}_t)$ is a strictly positive average of expected outcomes weighted by the conditional choice probability of each alternative. The function υ_{kt+1} in the recursive formulation of $A_t(h_t, \mathbb{E}_t)$ is a utility measure of income from occupation k received at the beginning of period $t+1$, adjusted for consumption smoothing. Higher values of the prior mean or higher values of human capital are associated with lower values of the index via the term υ_{kt+1}. Additionally, the value of human capital and beliefs in occupation k decreases with the size of the non-pecuniary costs, α_{kt}.

3.4.2 Occupation

Using equations (10) and (11), and applying logs to transform the problem, it can be shown that at any age t before retirement the individual chooses an occupation to solve

$$\max_k \sum_{k=0}^{4} d_{kt}\{\varepsilon_{kt} - \ln \alpha_{kt}(h_t) - (b_{\tau(t)} - 1) \ln E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})\upsilon_{kt+1}|\mathbb{E}_t, h_t]\} \qquad (12)$$

His occupational choice is independent of his current level of wealth. This is a consequence of the multiplicative separability of the ex-ante value function obtained in Proposition 1.[36] Trade-offs between occupations are characterized by differences in non-pecuniary utility components, ε_{kt} and $\alpha_{kt}(h_t)$, and differences in the expected utility from income, υ_{kt+1}, scaled by the index capturing the value of human capital accumulation and beliefs evolution, $A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})$. Since the taste shocks are assumed distributed Type-I Extreme Value, the expression in equation (12) becomes a standard logit. Following Hotz and Miller (1993), Proposition 2 uses the recursive nature of the index $A_{t+1}(h_t, \mathbb{E}_t)$ and yields a representation of the logarithm of the odds ratio in terms of future choice probabilities and utility

[36] In fact, relaxing the model to allow the use of savings as collateral would break this separability property as low levels of wealth will not allow the individual to obtain the optimal scale of his business, rendering the occupational choice dependent on wealth.

parameters.

Proposition 2. *For any choice $k > 0$, the logarithm of the likelihood ratio between choosing occupation k and choosing not to work is given by*

$$\ln\left(\frac{p_{kt}(h_t,\mathbb{E}_t)}{p_{0t}(h_t,\mathbb{E}_t)}\right) = -\ln\alpha_{kt}(h_t) - (b_{\tau(t)}-1)\ln E_t\left[v_{kt+1}\prod_{s=1}^{T-t}\left(\frac{p_{0t+s}(h_{kt}^{(s)},\mathbb{E}_{kt}^{(s)})}{p_{0t+s}(h_{0t}^{(s)},\mathbb{E}_{0t}^{(s)})}\right)^{\phi_t(s)}\middle|\,\mathbb{E}_t,h_t\right] \tag{13}$$

where

$$\phi_t(s) = \frac{1}{b_{\tau(t)+s}}\prod_{r=1}^{s-1}(1-1/b_{\tau(t)+r}) \tag{14}$$

and where $h_{kt}^{(s)}$ and $\mathbb{E}_{kt}^{(s)}$ indicate the value of the state variables at future age $t+s$, conditional on the decision path described by making $d=1$ for all $d \in \{d_{kt}, d_{0t+1}, d_{0t+2}, \ldots, d_{0T}\}$.

Proof: See Model Appendix.

Equation (13) shows that the logarithm of the likelihood ratio between working in any occupation $k > 0$ and the decision not to work is a function of the trade-offs described in equation (12).[37] Higher non-pecuniary costs and lower expected utility from compensation make alternative k less likely. Moreover, if choosing alternative k makes the individual less likely to work in the future, thereby reducing the value of his human capital or his information, then alternative k is also less likely to be chosen today.

4 Estimation

Estimation of the parameters of the model is done in two stages using a combination of an Expectation-Maximization (EM) algorithm and a conditional choice probabilities (ccp) estimator (Hotz and Miller (1993), Arcidiacono and Miller (2011), James (2011)). The EM algorithm in the first stage permits fast estimation of income and ability parameters, bypassing the need for multidimensional integration over unobserved ability vectors. The ccp estimator in the second stage follows naturally from the expression derived in Proposition 2. It allows for a flexible treatment of the large state space of the problem, which includes continuous beliefs and experience. It yields consistent estimates of the utility parameters without having to solve the dynamic optimization problem at every candidate parameter

[37]The derivation of equation (13) relies on the assumption that the decision not to work changes neither the beliefs nor the vector of accumulated human capital. In general, an expression similar to that in Proposition 2 can be obtained in terms of any path of future choices. Such alternative representation, however, will not be as tractable because expectations over functions of future income signals will be necessary.

vector in the search algorithm. The flexibility provided by the ccp estimator facilitates the disaggregation of occupations used in this paper.

Individuals in the model select on beliefs about their ability rather than selecting on their actual ability. In other words, conditional on the history of income signals up to t, mapped into beliefs $\mathbb{E}_{it}$, choices at t are independent of ability. Let Λ be the collection of parameters of the utility function, let Θ be the collection of income parameters, including the variance parameters of the productivity shocks, σ_{η_k}, and let $\mathbf{\Delta}_s$ be the covariance matrix of the population ability distribution conditional on education level s. Therefore, the likelihood of the data—hourly income and choices—for a person with education level s can be written as:

$$\mathcal{L}_i = \prod_{t=t_{i0}}^{T_i} \prod_{k=0}^{4} \Pr\left[d_{kit} = 1 | h_{it}, \mathbb{E}_{it}; \Lambda, \Theta\right]^{d_{kit}} \times \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_{i0}}^{T_i} \prod_{j=1}^{4} \Pr\left[y_{jit+1} | h_{it}, \tilde{\mu}_j; \Theta\right]^{d_{jit}} \right\} dF\left(\tilde{\mathcal{M}}; \mathbf{\Delta}_s\right) \tag{15}$$

Equation (15) reflects two characteristics of the model. First, individuals are heterogeneous in their unobserved ability, $\mathcal{M}_i$. Second, rather than their unobserved ability, a function of the history of their income signals—their belief $\mathbb{E}_{it}$—shapes their occupational decisions. This has the convenient effect of taking the choices part of the likelihood out of the multi-dimensional integral.[38] The log likelihood of (15) is then additively separable:

$$\ln \mathcal{L}_i = \ln \mathcal{L}_i^d + \ln \mathcal{L}_i^y$$

The first stage of the estimation procedure utilizes the income term of the log likelihood to obtain estimates of Θ and $\mathbf{\Delta}_s$. These estimates are used in the second stage to estimate Λ. The scale of Θ, $\mathbf{\Delta}_s$, and ρ depends on the units in which income and consumption are measured. Hourly income is expressed in \$10 units and consumption in \$1,000 units. Therefore, converting hourly income into annual income for occupation k in the model entails dividing $\bar{L}_k y_{kt+1}$ by 100. The standard errors provided in the current version of the paper are uncorrected for the two-stage estimation.[39]

[38] No measurement error is assumed in the hourly income data, which allows for people's beliefs to be backed out using income data. Allowing for measurement error would render the two-stage procedure non-viable as integration over the error terms would be necessary over the entire expression in equation (15).

[39] For the parameters in the first stage, which uses an EM algorithm, computation of the standard errors follows the NDS procedure described in Jamshidian and Jennrich (2000).

4.1 First Stage: Income Parameters and Learning Structure

The first stage uses an Expectation-Maximization (EM) algorithm, an iterative method that yields maximum likelihood estimates when a portion of the data is unobserved. In the model, the unobserved part of the data is the individual's ability, $\mathcal{M}_i$. In order to implement the EM algorithm, assume $\mathcal{M}_i$ is observed for all i. Hence, the income term of the log likelihood for individual i becomes

$$\ln \mathcal{L}_i^y(\mathcal{M}_i) = \sum_{t=t_{i0}}^{T_i} \sum_{j=0}^{4} d_{jit} \ln \Pr\left[y_{jit+1} | h_{it}, \mu_j; \Theta\right]$$

Starting from a guess of parameters $\langle \Theta^0, \Delta^0 \rangle$, implementation of the EM algorithm entails iteration over the following two steps to obtain maximum likelihood estimates:

1. *Expectation Step*: compute the expected value of $\ln \mathcal{L}_i^y(\mathcal{M}_i)$, conditional on the data actually observed and the parameters at the mth iteration

$$E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i)|\cdot]$$

2. *Maximization Step*: find the new iterated value of the vector of parameters by maximizing the expression obtained in the expectation step:

$$\langle \Theta^{m+1}, \Delta^{m+1} \rangle = \max_{\langle \Theta, \Delta \rangle} \sum_i E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i)|\cdot]$$

For use in the second stage, consistent estimates of individual beliefs are obtained. This computation uses the point estimates for Θ and $\boldsymbol{\Delta}_s$, the history of signals received by every individual, Bayes' Rule, and the rational expectations assumption regarding the individual's prior. The first stage of the estimation algorithm is detailed in the Estimation Appendix.

4.2 Second Stage: Utility Parameters

In order to estimate Λ, the second stage follows Hotz and Miller (1993) and takes advantage of the expression derived in Proposition 2 mapping future choice probabilities and utility parameters into current choice probabilities. The Type-I Extreme Value assumption regarding the distribution of preference shocks implies that the choice probabilities can be written as

$$p_{kit}(h_{it}, \mathbb{E}_{it}) = \frac{\exp(V_k(h_{it}, \mathbb{E}_{it}))}{1 + \sum_{k'>0} \exp(V_{k'}(h_{it}, \mathbb{E}_{it}))} \tag{16}$$

where $V_0 = 0$ and for any $k > 0$

$$V_k(h_{it}, \mathbb{E}_{it}) = -\ln \alpha_{kit}(h_{it}) - (b_{\tau(t)} - 1) \ln E_t \left[v_{kit+1} \prod_{s=1}^{T-t} \left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{0it}^{(s)})} \right)^{\phi_t(s)} \middle| \mathbb{E}_{it}, h_{it} \right] \tag{17}$$

An iterative algorithm is implemented that maximizes the log likelihood of the data while searching over the space of parameters and ccps. This procedure is akin to the swapping of the nested fixed point algorithm described in Aguirregabiria and Mira (2002). The procedure is initialized with flexible parametric versions of the future conditional choice probabilities estimated from the data.[40] It entails the following two steps:

1. *Maximization Step*: plug the estimated ccps in equation (16) and maximize the log likelihood of the observed choices. Notably, forming $V_k(h_t, \mathbb{E}_t)$ requires knowledge of the bond prices $b_{\tau(t)}$. Following Gayle and Miller (2009), obtain bond prices using series from the Federal Reserve Economic Data (see Data Appendix).

2. *CCP Step*: use the estimated parameters at the current iteration to solve the model backwards and obtain new ccps implied by the model.

The parameter vector that yields the minimum log likelihood is chosen. The second stage of the estimation method is further explained in the Estimation Appendix.

4.3 Identification

The sources of variation that identify the parameters of the model are observed income and occupational choices over time, as well as variation in observed experience and demographics. There are three sets of parameters to be identified: returns to experience and education, parameters of the Bayesian learning structure including the distribution of abilities, and parameters of the utility function. Identification of each set of parameters is discussed bellow.

The main challenge for causal estimation of the returns to occupation-specific experience in the income equation is that individuals select on beliefs (Gibbons et al. (2005)). This is accounted for in estimation using the likelihood induced by the model.[41] Provided that selection is accounted for, income variation for different levels of occupation-specific experience identifies the returns to experience. This is the learning-by-doing piece of the model. The

[40]When computing the ccps, beliefs, estimated in the first stage, are also treated as data.

[41]Instrumental variables have also been used to account for selection on beliefs (see Altonji and Williams (2005), Dillon and Stanton (2016)).

returns to cross-occupation experience are identified from the variation in income of switchers. Notably, since education indicators are introduced linearly in the income equations, the mean of the distribution of ability is not identified.[42]

Identification of the latent distribution of ability under normality assumptions using income data and occupation choices has been shown in Heckman and Honoré (1990). Individuals with different education levels are likely to have different distributions of ability. Therefore, given that education is not endogenized, the distribution of ability is made education-specific. The parameters of the covariance matrix of the ability distribution are then identified from variation in residual income. In particular, the off-diagonal terms of the variance matrix are identified from the covariation in residuals of switchers. Finally, the variance of the idiosyncratic shocks, which is not made education-specific, is identified from the excess variation in residual income per occupation.[43]

Following results in Arcidiacono and Miller (2015), given that the distribution of the choice-specific taste shocks, the subjective discount factor, and the transition function of beliefs—estimated in the first stage—are known, the flow payoffs are identified up to the normalization that the flow payoffs from unemployment are zero in each state and time period.[44] Hence, the functional form assumptions regarding the utility function provide over-identifying restrictions. Notably, separate identification of the risk aversion parameter from prior beliefs may be difficult. This is because similar choice patterns would be generated by overconfident low-variance priors and high risk aversion, or by under confident high-variance priors with low risk aversion. The panel dimension of the data helps separate the variation in choices due to risk aversion. Over time and regardless of priors, Bayesian learning implies that individuals' beliefs will get arbitrarily close to their true ability, and the remaining idiosyncratic variation would help identify the risk aversion parameter.

5 Parameter Estimates and Economic Forces at Play

In this section, the estimates of the structural model are discussed and the economic forces at play in the decision to become an entrepreneur are evaluated. Especial attention is paid to the role that these forces play in the timing of the entrepreneurial choice. Section 5.1 presents estimated parameters of the income equation and the distribution of ability. Section 5.2

[42] The distribution of ability is also conditional on education and its mean enters linearly in the income equation.

[43] These points can be seen more clearly in the updating rules of the EM algorithm in the Estimation Appendix.

[44] Identification in models of dynamic discrete choice is also discussed in Magnac and Thesmar (2002) and Rust (1994).

discusses utility parameter estimates and Section 5.3 discusses model fit. Finally, the role of the forces at play is quantified using a decomposition exercise in Section 5.4.

5.1 Income Parameters and the Distribution of Ability

Two key economic forces in the model are learning-by-doing and learning about ability. In this section the estimated results for these processes are described. Beginning with learning-by-doing, Table 6 presents the estimated returns to experience as well as other parameters of the income equation in (4). The returns to experience are specified using step-functions that capture the marginal increase in productivity for an extra year of experience.[45]

The estimated increments in productivity from Table 6 are illustrated in Figure 4. As an example, one year of blue collar experience adds .81 dollars to the hourly income from blue collar work. A second year of blue collar experience adds another .74 dollars and so forth. Figure 4 shows that the returns to blue collar experience are the flattest while the returns to incorporated experience are the steepest. Individuals trying to reach the high productivity levels available in incorporated entrepreneurship should start climbing the ladder when they are young.

Learning-by-doing can also happen across-occupations. This is, experience accumulated in one occupation may have non-zero returns in another. Figure 5 illustrates the estimated returns to cross-occupation experience. As an example, one year of white collar experience adds 1.33 dollars to the hourly income from blue collar work, and any extra year of white collar experience up to five does not add anything. The most striking result coming out of Figure 5 is that, while expertise in entrepreneurial activities always increases productivity in paid employment, low levels of entrepreneurial experience reduce it. This finding is similar to some of the results in Jovanovic and Nyarko (1996), where switching technologies can reduce productivity by reducing expertise.[46] These results suggest that, unless mastered, skills learned in entrepreneurship may harm paid-employment productivity.[47]

Another form of learning allowed in the model is learning about ability. In the model, individuals start their careers with a common prior that corresponds to the population

[45]The choice of representing the returns profile using a step-function is data driven. It is difficult to obtain smooth profiles for entrepreneurial experience given that the number of individuals with high numbers of entrepreneurial experience is low. The location of the steps was chosen from preliminary OLS regressions. No steps beyond the 10th year of experience were significant in the OLS exercise so it is assumed that individuals reach the top of the productivity ladder by the 10th year in the occupation.

[46]Manso (2014) finds that entrepreneurial experience generates a premium for salaried workers whenever it is more than two years.

[47]Some skills commonly considered entrepreneurial are the ability to sell, innovative thinking, initiative, and self-reliance.

distribution of ability for their education level. Over time, their success or failure in each occupation allow them to update their beliefs. The estimated covariance matrices for the ability distribution, which constitute the initial priors, are presented in Table 7.[48] Recall that the ability vector is measured in \$10s/hr. Therefore, an individual with more than college education and ability that is one standard deviation above the population mean would have an incorporated ability of $10 \times \sqrt{10.88} \approx \$33$/hr. The same individual, would have white collar ability of $10 \times \sqrt{0.87} \approx \$9$/hr. This comparison highlights the most important results from Table 7: there is higher variation in entrepreneurial ability than in paid-employment ability. In addition, more education tends to be associated with higher variation in ability. In light of the model, learning about ability implies that individuals should enter entrepreneurship as early as possible in order to find out their position in the high-variance ability distribution.

Learning about ability can also happen across-occupations because occupational abilities are correlated. This can be seen in the off-diagonal terms of the covariance matrix in Table 7. To better understand this relationship, the correlations between abilities are computed and presented in Figure 6. Individuals with high ability in white collar work tend to have high ability in entrepreneurial activities. Moreover, consistent with the discussion in the Section 2, the correlation between white collar ability and incorporated ability is higher than the correlation between both entrepreneurial abilities at any education level. These results are also consistent with the differences between entrepreneurial occupations presented in Levine and Rubinstein (Forthcoming).

Estimates indicate that there is an incentive for young individuals to attempt entrepreneurship in order to learn whether they are high-ability. Alternatively, they can use white collar success as an indicator of their entrepreneurial ability. However, these learning possibilities depend on how noisy the signals are in each occupation. As Table 8 shows, the high idiosyncratic variation of entrepreneurial occupations threatens their informational value.

To get a sense of how fast own- and cross-occupation learning about ability can actually happen, Figure 7 presents the percent of prior uncertainty about entrepreneurial ability that is eliminated after working for 5 years in each occupation. Consider the right panel and individuals with more than college education—the rightmost set of 4 bars in the figure. On the one hand, own-occupation learning indicates that initial uncertainty about incorporated ability is reduced by almost 90 percent, after 5 years of incorporated experience. On the other, cross-occupation learning indicates that initial uncertainty about incorporated ability is reduced by about 30 percent, after 5 years of white collar experience. Surprisingly, Figure

[48] As mentioned above, the mean of the ability distribution is normalized to zero because it is not separately identified from the linear returns to education in equation (4) (see Table 6).

7 shows that for college educated individuals, paid employment can be a better source for learning about incorporated ability than incorporated entrepreneurship itself. Correlated learning about ability provides incentives for young college educated individuals to attempt entrepreneurship later in life.

5.2 Utility Parameters

Estimates of the utility function parameters are presented in Table 9. First, individuals are risk-averse—the point estimate for ρ is greater than zero. To better assess the importance of risk aversion, Figure 8 depicts static and dynamic measures of certainty equivalence. On the y-axis of both panels is the proportion of expected annual income that would be necessary to make individuals indifferent between accepting the income variation in the occupation and receiving the certainty equivalent for sure.[49] The static certainty equivalent, on the left panel in Figure 8, highlights the role of risk aversion as it does not account for the future value of experience and beliefs. Not surprisingly, incorporated entrepreneurship has the lowest static certainty equivalent as it offers the highest income variation. However, dynamic considerations about future human capital and beliefs mitigate the effects of risk aversion (right panel of Figure 8).

Besides risk aversion, first-time entry costs are a barrier to young entrepreneurship. Table 9 suggest that entry costs for entrepreneurship are higher for young individuals as well as individuals with low permanent wealth. Figure 9 introduces monetary equivalents for entry cost estimates.[50] On the x-axis of each panel is the education level and on the y-axis is the monetary equivalent in thousands of dollars (year, 2000). Entry costs are more responsive to age than to permanent wealth. On the one hand, the monetary equivalent of entry costs into incorporated entrepreneurship decreases from about \$200,000, for individuals age 20, to less than \$150,000, for individuals age 40. On the other, entrepreneurial entry costs decrease \$20,000 or less when permanent wealth goes from the 10th percentile to the 90th. The negative sign of the relation between entry costs and permanent wealth suggests, in a reduced form sense, individuals are able to ease their barriers to entrepreneurship using their life-time potential. However, the steeper profile of entry costs associated with age captures barriers to entry not explicitly modeled, such as tighter credit constraints for young individuals with weaker credit histories or less capital.[51]

[49] By construction, provided individuals are risk-averse, this measure is always bounded above by 1 in the static case. The measure is further explained in the Results Appendix.

[50] In the specification of the model, entry costs and non-pecuniary benefits can be treated as the indirect utility representation of terms in the budget constraints (see Results Appendix).

[51] One more pattern emerges from the figure. the dispersion in entry costs across-occupations decreases

Non-pecuniary motivations such as "being one's own boss" and "wanting flexibility over schedule" (Hamilton (2000), Hurst and Pugsley (2011), Hurst and Pugsley (2015)) have also been suggested in the literature. This association would be even stronger for risk-averse individuals trying to avoid the higher variation of entrepreneurial outcomes. However, the dynamic treatment of the entrepreneurial choice, as well as the integration of its information value, suggests a more nuanced story. Estimates of the non-pecuniary benefits not associated with first entry are presented in Table 9 and are converted to their monetary equivalent in Figure 10. Overall dominance of entrepreneurial non-pecuniary benefits does not emerge once dynamic considerations are introduced.[52] On the one hand, entrepreneurial activities are always ranked below blue collar work for low educated individuals. In monetary terms, this difference is equivalent to at least $20,000 per year. On the other, entrepreneurial occupations become more attractive in non-pecuniary benefits for individuals with college or more. In particular, the non-pecuniary benefits of incorporated entrepreneurship for the college educated are higher that those from any other occupation at any education level. Hence, it is possible that the importance of non-pecuniary benefits explaining the entrepreneurial choice has an education gradient not explored in previous literature.

Finally, even though the treatment of the entrepreneurial choice in this paper attempts to capture many of its economic determinants, there are at least two caveats of this analysis. First, credit constraints are not modeled explicitly. Data limitations as well as model tractability kept this project away from incorporating the role of savings in entrepreneurship. Credit constraints remain as a common explanation for the lack of higher entrepreneurial participation (Evans and Jovanovic (1989), Hurst and Lusardi (2004), and Buera (2009)). In reduced form, the age profile of entry costs in the model tries to get at this issue. Consistent with the credit constraints hypothesis, results show young individuals face higher entry costs to entrepreneurship. Second, the decision of hours worked is not modeled. Hence, the non-pecuniary benefits from working in a given occupation are net of the disutility from working. Given that entrepreneurs work more hours, it is possible that accounting for the disutility from working would yield non-pecuniary benefits that are higher for entrepreneurial occupations, at least for the highly educated.

with age as entrepreneurial entry costs decline but salaried entry costs increase. This phenomenon may result from older individuals finding it harder to start careers in paid employment due to difficulties in obtaining or regaining skills at old ages.

[52] These results are in line with those in Dillon and Stanton (2016), who account for dynamics highlighting the option value of the entrepreneurial choice.

5.3 Model Fit

In order to assess goodness of fit, the model is solved and data are simulated (see Results Appendix). Simulated choice rates contrasted against the data in Figure 11 show that the model successfully generates the incorporated and unincorporated participation rates over the life cycle. However, the model under-predicts white collar participation and over-predicts blue collar participation.[53]

In order to investigate whether the model replicates the absence of young entrepreneurs, first-entry statistics are presented in Table 10. The model captures well the proportion of individuals who attempt entrepreneurial occupations by age 40. It also captures reasonably well the average age at first entry into all occupations. More interestingly, the model captures the nature of the experience obtained before first entry. Consistent with the similarities between white collar work and incorporated entrepreneurship highlighted in Section 2, simulated individuals attempting entrepreneurship tend to have more prior experience. In particular, first-time unincorporated entrepreneurs tend to have more blue collar experience, whereas the opposite is true for first-time incorporated entrepreneurs. Further measures of model fit in terms of transitions, spells, and realized income are presented in the Results Appendix.

5.4 Decomposition Exercises

In this section, the effects of the economic forces at play are quantified by comparing simulated data from the estimated model (baseline) against simulated data from a number of counterfactual regimes that disable parts of the structure. The quantification is done in four dimensions: entry, timing, ability, and present value of income. Given that incorporated entrepreneurs seem most comparable to what is commonly considered as "the entrepreneur" (see Section 2 as well as Levine and Rubinstein (Forthcoming)), the discussion will include both types but will center on incorporated entrepreneurs. Comparisons reveal the two main barriers to young entrepreneurship are entry costs and lack of information. Moreover, information frictions have a large long-term effect: fully-informed incorporated entrepreneurs have a present value of income (PVI) that is about 50% higher than in the baseline. Additionally, the decomposition points to a considerable long-term effect from using paid-employment outcomes to predict incorporated entrepreneurial success.

[53] In a separate exercise, not shown here for space considerations, choice rates are simulated at any age t taking the state observed in the data as given. This exercise shows a much better fit for salaried occupations suggesting that the model has a harder time capturing transitions.

The counterfactual regimes used for decomposition are:[54]

- ↪ *No leaning-by-doing*: productivity does not increase with own or cross-occupation experience. Instead, occupational skill is constant and pays an average return.[55]
- ↪ *No learning about ability*: individuals know their ability vector $\mathcal{M}_i$ but the initial level of uncertainty remains unchanged.[56]
- ↪ *No X (cross-occupation) leaning-by-doing*: productivity in one occupation is invariant to experience in another. The cross-occupation returns to experience (see Figure 5) are set to zero.
- ↪ *No X (correlated) learning about ability*: individuals believe that their success in one occupation is uninformative of their ability in another. Their initial prior variance is $\mathbf{\Delta}_s$ diagonalized.
- ↪ *No uncertainty*: individuals know their ability vector $\mathcal{M}_i$ and there is no idiosyncratic variation around their hourly income.
- ↪ *Uniform entry cost*: entry costs does not vary with age. Instead, individuals pay the cost faced by a 35-year old individual with their same education level.

Why are there not more entrepreneurs? Having to climb the productivity ladder has the strongest effect discouraging incorporated entry. Close behind are the effects of uncertainty (in ability and idiosyncratic variation) and entry costs. Figure 12 displays the ratio of the share of individuals who attempt entrepreneurship during their careers in each of the counterfactuals relative to the baseline. It shows that shutting down learning-by-doing has the strongest effect on incorporated entry: if people did not have to learn-by-doing their way up through the steep incorporated productivity ladder (see Figure 4), they would be almost twice as likely to attempt incorporated entrepreneurship. In isolation, shutting down learning

[54]The counterfactual regimes are further explained in the Appendix. The initial state is fixed across counterfactual regimes. Extended results from these comparisons are presented in the Results Appendix.

[55]Let $R_k(x)$ be the return to experience in occupation k for somebody who has worked x years in occupation k and zero years in any other occupation (see Figure 4). The fixed hourly return to observed skill for individuals in occupation k under this counterfactual regime is

$$\bar{y}_k = \sum_{x=0}^{20} R_k(x)$$

[56]In terms of equation (4), this amounts to changing the value of the idiosyncratic income variance in occupation k from just σ_{η_k} to $\sigma_{\eta_k} + \mathbf{\Delta}_{s,\{k,k\}}$.

about ability (i.e. sorting on ability) only increases the rate of incorporated entrepreneurship by about 35% over the baseline.[57] However, once all uncertainty is eliminated, the effect on entry is similar to the effect from flattening the entry cost: individuals would be 75% more likely to attempt incorporated entrepreneurship.[58] Shutting down either type of cross-occupation learning decreases incorporated participation. This latter result makes sense because white collar workers, who are set to gain the most from switching, cannot increase their productivity or improve their beliefs using their experience and paid-employment success.

What about young entrepreneurship? The two main barriers to young entrepreneurship are entry costs and lack of information. To see this, the gap in first entry age between entrepreneurial occupations and white collar work, introduced in Table 4, is analyzed. Figure 13 shows the percentage of that gap that is closed under each of the counterfactual regimes. Flattening the entry costs closes 70% of the gap and sorting on ability closes 20%. If individuals knew their ability more would enter early. Risk aversion and correlated learning do indeed induce young individuals to bypass some of the risk of attempting entrepreneurship by acquiring some paid-employment experience first. Figure 13 shows that eliminating all uncertainty reduces the gap by an extra 5% on top of the reduction attained from providing full information about ability. Interestingly, eliminating correlated learning on its own widens the gap by about 3%. In the uncorrelated learning case, this happens because young individuals who avoid starting their careers as entrepreneurs must first find out that they really are not good at paid employment in order to switch into entrepreneurship.

However, individual participation is not all that matters. The ability of those attempting entrepreneurship can have strong effects for the economy as a whole. Figure 14 displays the ratio of ability at first entry in the counterfactual regime relative to the baseline. Not surprisingly, the ability of fully-informed individuals entering incorporated entrepreneurship for the first time is higher—about three times as large as in the baseline. Perhaps more surprising is the fact that the ability of fully-informed individuals entering unincorporated entrepreneurship is twelve times as large as in the baseline. This suggests that under full information the returns from unincorporated entrepreneurship are relatively less attractive. Hence, individuals choosing to enter must be of very high unincorporated ability. Shutting down learning-by-doing and flattening entry costs reduce incorporated ability at first entry relative to the baseline because they ease the threshold to enter. Interestingly, shutting down correlated learning not only reduces ability at first entry but makes it negative. The

[57] The role of information frictions in a correlated learning framework has also been studied empirically in schooling and work decisions Arcidiacono et al. (2016)

[58] Given the results in Table 9, flattening the entry costs with respect to age amounts to young people facing a lower entry cost than in the baseline.

reason for this is the mismatch between prior variance and population variance. In the counterfactual regime, although individuals only learn about ability if they attempt the occupation, the true ability distributions are still characterized by the covariance matrices in Table 7. Provided abilities are mostly positively correlated (see Figure 6), this means that those who switch to entrepreneurship are likely to be low ability entrepreneurs. After all, they switch after discovering their low paid-employment ability.

Finally, the economic forces in the model have long-term effects on the outcomes of entrepreneurs. Figure 15 shows the present value of the entrepreneurs' realized stream of income relative to the baseline. Results indicate that the effect of the information frictions is large: fully-informed incorporated entrepreneurs have a present value of income (PVI) that is about 50% higher than in the baseline. Flattening entry costs also increases the PVI of the incorporated, although only by about 10%, because successful young entrepreneurs will enjoy the returns for longer. Shutting down learning-by-doing or shutting down cross-occupation learning-by-doing both reduce the PVI by about 20%. Since the former is an extension of the latter, what really decreases the incorporated PVI is the lack transferability of skills learned. Notably, shutting down correlated learning decreases the incorporated PVI by about 25%. The long-term effect of using paid-employment outcomes to predict incorporated entrepreneurial success is certainly not negligible.[59]

6 Policy Counterfactuals

According to the decomposition exercises in Section 5, the two main barriers to young entrepreneurship are entry costs and information frictions. In this section, policies focusing on incorporated entrepreneurship target these barriers. This section extends the literature by providing a mapping from entrepreneurial education that shifts beliefs, into career choices and long-term outcomes. Results suggest that a blanket subsidy increases entrepreneurship but has small long-term effect as measured by the PVI. Additionally, results show that entrepreneurial education that provides information can have sizable effects on participation and present value of income flows, even for low information quality.

6.1 Subsidies

The previous section showed that entry costs are a strong barrier to young entrepreneurship. Consequently, the intervention considered here is a subsidy for young incorporated

[59]Simulations in the Results Appendix show that a model with uncorrelated learning does not reproduce the trends observed in Figure 2. The baseline model with correlated learning does reproduce the trends.

entrepreneurs. The policy consists of giving either \$25,000 or \$50,000 to any individual who decides to start his career as a young incorporated entrepreneur. This is, the subsidy is granted only if the individual becomes an entrepreneurs immediately after finishing his education.[60] Table 11 summarizes the effects of the intervention in terms of young incorporated entrepreneurs and in terms of the overall pool of incorporated entrepreneurs (long-term effects).

The results reflect the effects of lowering the threshold of entry when information is imperfect. There is more participation but less entrepreneurial quality. The \$50,000 subsidy more than doubles young incorporated entrepreneurship as measured by the number of individuals who attempt entrepreneurship during their first five years in the labor market. However, the average ability at first entry decreases by about 60%.

Regardless of this adverse effect on the quality of young incorporated entrepreneurs, the subsidy may be justified. For instance, the subsidy may attract high-ability individuals who were not entering due to high entry costs associated to their age or lower permanent wealth. However, as measured by the 95th percentile of the ability of young entrepreneurs, this is not the case. The subsidy could also have long-term effects. As a consequence of the \$50,000 subsidy, the number of people who attempt incorporated entrepreneurship in their careers increases by 50%. Besides, the average net present value of income of all individuals in the economy (those who experiment with entrepreneurship and those who do not) increases by about 3%. These results suggest that the marginal individuals induced to experiment by the subsidy turn out to be more productive as entrepreneurs than what they would have been as paid employees.

Looking at the average entrepreneurial ability of new entrants, the results here seem consistent with those in Hamilton et al. (2016) and underscore the arguments presented in Shane (2009) against blindly subsidizing entrepreneurship. However, once the long-term effects are evaluated, it appears that policies that relax entry costs and attract marginal entrepreneurs may be effective. Although small, there are gains in terms of PVI from young successful entrepreneurs attracted early on by the subsidy.

[60] One such example of interventions providing funds for young entrepreneurs is the Thiel Fellowship (`http://thielfellowship.org`). However, this intervention relies on a tournament that may reveal information about the participant's quality.

6.2 Entrepreneurship Education

Many policies trying to foster young entrepreneurship focus on entrepreneurship education.[61] To the extent that these policies help reveal entrepreneurial potential, the emphasis on entrepreneurial education is consistent with the decomposition exercises in Section 5 showing that information frictions are a main barrier to young entrepreneurs. The empirical literature has provided evidence that entrepreneurship education programs can shift individuals' elicited beliefs and intentions (von Graevenitz et al. (2010), Oosterbeek et al. (2010), Souitaris et al. (2007), Peterman and Kennedy (2003)), but the value of these policies will critically depend on the quality of the information they provide. Results here extend the literature providing a mapping from entrepreneurship education of a given quality, through shifts in beliefs, into career choices and long-term outcomes. In other words, the paper provides a dynamic framework to assess the value of any given policy that provides information.[62]

The counterfactuals operate as follows: all individuals draw noisy information—regarding their ability as incorporated entrepreneurs—from their outcomes in an entrepreneurship education program. Individuals use this information to update their beliefs before beginning their careers. This policy effectively induces initial heterogeneity in entrepreneurial beliefs that will depend on ability as well as on luck. Additionally, because abilities are correlated, the policy will induce heterogeneity in beliefs across all occupations. In the language of the model, the entrepreneurship education program yields every individual a signal about his incorporated ability ($\mu_{4,i} \in \mathcal{M}_i$) given by

$$\zeta_i^p = \mu_{4,i} + \nu_i \tag{18}$$

where ν_i are iid $N(0, \sigma_\nu^2)$. Individuals use the information contained in ζ_i^p to update their beliefs before entering the labor market. It is assumed that no entrepreneurship education program can provide better information than actually becoming an entrepreneur in the job market for one period. In other words, the noise variance from this intervention is bounded below by the idiosyncratic variance ($\sigma_{\eta_4}^2$) estimated in Section 5 (see Table 8). Therefore, the noise variance from the entrepreneurship education program can be written as

$$\sigma_\nu^2 = s \cdot \sigma_{\eta_4}^2, \quad \text{with } s \geq 1 \tag{19}$$

[61] Examples of such programs are the BizCamps or the Regional Young Entrepreneurship Challenge by the Network for Teaching Entrepreneurship and the Junior Achievement Young Enterprise Student Mini-Company (SMC) program.

[62] The paper remains silent as to what the information quality of entrepreneurship education is. Evaluation of the information quality of a specific policy could be attained from observing the information signal generated through the program (for instance, rankings, scores, simulated profits) to randomly selected individuals.

Four counterfactuals are considered here that differ in terms of how noisy their signals are: $s \in \{1, 2, 5, 10\}$. For instance, when s equals 10 the quality of information from the entrepreneurship education program is 10% the quality of information from actually becoming an incorporated entrepreneur for one period. Results from these interventions are presented in Table 12.

For the range of information quality considered ($s \in \{1, 2, 5, 10\}$), results suggest that the lower the quality, the higher the percentage of young entrepreneurs and the bias in beliefs that induces them to enter. Table 12 also shows a decline in average ability of young entrepreneurs for information quality below 50% ($s = 2$). This reflects the amount of young entrepreneurs who are attracted by lucky signals in programs with lower information quality. Providing noisy information magnifies the role of overestimation of ability in fostering experimentation. Young incorporated entrepreneurs go from having an hourly-income negative bias of -\$1.3/hr in the baseline to a positive bias of \$81/hr from entrepreneurial education providing 10% information quality ($s = 10$). These results are consistent with previous literature suggesting that overconfidence influences entrepreneurial entry (Roll (1986), Camerer and Lovallo (1999)). However, in the framework of this paper, overestimation of ability is not a different psychological trait of entrepreneurs or the result of differential analysis of the information received (March and Shapira (1987), Busenitz (1999)). Instead, overestimation at first entry emerges endogenously from uninformed rational individuals who are fortunate to receive large positive signals.

Entrepreneurial education can also provide long-term gains even when the quality of information is as low as 10%. Table 12 shows that the share of incorporated entrepreneurship at age 40 triples, the percentage of individuals who attempt incorporated entrepreneurship increases by about 70%, and the PVI of incorporated entrepreneurs increases by 25%. Additionally, entrepreneurial education could benefit all individuals, not only those who eventually become entrepreneurs. A first approximation to the benefits from entrepreneurial education to the average individual is given by the difference in average PVI relative to the baseline. According to the last row in Table 12, this number varies from \$65,000 when $s = 10$ to \$73,000 when $s = 1$.[63]

The results in this section extend the literature providing a mapping from movements in beliefs, generated by entrepreneurial education of a given quality, into career choices and long-term outcomes. Nevertheless, caution must be taken when reading these results. The

[63] To reflect the differences in entrepreneurial potential at every education level, these quantity can also be computed by education level. Results imply that the difference in expected PVI from entrepreneurial education of 10% of quality goes from \$700 for high schoolers to about \$200,000 for individuals with more than college education.

information quality of any specific policy may be different and its cost may well exceed the additional income it generates. In the limit, as s goes to infinity, entrepreneurship education loses all its information value.[64]

7 Conclusion

On the basis of their potential economic benefits, entrepreneurship, and in particular young entrepreneurship, are the target of many policy interventions. These interventions are a response to the fact that most individuals do not start a business and, if they do, tend do so well into their thirties after they have accumulated some salaried experience. While interventions aiming to foster young entrepreneurship spread, their effectiveness is still an open question.

This paper explores the reasons why individuals attempt entrepreneurship in their careers as well as the reasons explaining the gap in first-entry ages between paid employment and self-employment. The paper extends the literature by quantifying the relative importance of various determinants of entrepreneurial choice studied separately in previous research, namely, accumulation of experience (learning-by-doing), accumulation of information (learning about ability), risk aversion, and entry costs. In addition to the quantification, three elements are particularly novel to the empirical literature on entrepreneurship: an analysis of the gap in first-entry ages, an assessment of the relative importance of risk aversion in a dynamic setting, and an evaluation of the role of cross-occupation learning between paid employment and entrepreneurship.

Using the structure of the model to quantify effects, a decomposition exercise indicates that learning-by-doing and entry costs have the largest effects preventing individuals from attempting entrepreneurship. Risk aversion and information frictions also play important roles. For instance, shutting down risk aversion increases the percentage of individuals who attempt incorporated entrepreneurship by 40% and eliminating information frictions increases this number by 35%. Eliminating cross-occupation learning reduces by 10% the percentage of individuals attempting entrepreneurship. Although these effects may seem small, in the long term the effects become stronger: eliminating cross-occupation learning decreases the present value of income of incorporated entrepreneurs by about 25%.

[64] A more general characterization of the information value of entrepreneurship education programs could be to write

$$\zeta_i^p = \kappa \cdot \mu_{4,i} + \nu_i$$

for $\kappa \in [0, 1]$. An uninformative entrepreneurial education program could also be one in which entrepreneurial ability is only weakly related to outcomes, i.e. where κ approaches zero.

Results also indicate that the main determinants of the gap in first entry ages are entry costs and information frictions. The paper evaluates the effects of policies that target these barriers. Previous literature has shown that entrepreneurship education can shift beliefs (von Graevenitz et al. (2010), Oosterbeek et al. (2010)). This paper extends the literature by providing a mapping from the information quality of entrepreneurship education into career choices and outcomes. Results show that a blanket subsidy increases young entrepreneurship by lowering the threshold of entry but has a limited long-term effect. Additionally, results indicate that entrepreneurship education can have large effects fostering young entrepreneurship and increasing the present value of income for all individuals. Nevertheless, the information content of any specific policy must be assessed separately. It is possible for some policies to be very uninformative, compromising the benefits from implementing them.

Finally, the motivation for many entrepreneurship policies goes beyond a simple desire to attract more entrepreneurs. Many policies seeking to foster entrepreneurship are motivated by the jobs entrepreneurs create. However, beyond attracting new entrants, entrepreneurship policies may or may not affect the decision to hire employees (Fairlie and Miranda (Forthcoming)). Therefore, more work is needed to evaluate these policies taking into account not only the amount of new entrepreneurs but also their quality and their propensity to generate jobs. The framework introduced here is the first step towards that goal. Future research could account for the effects of entrepreneurship policies on job creation by extending the model and acquiring data on the number of employees hired.

References

Aguirregabiria, V. and P. Mira, "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 2002, *70* (4), 1519–1543.

Altonji, Joseph G. and Nicolas Williams, "Do Wages Rise with Seniority? A Reassessment," *ILR Review*, 2005, *58* (2), 370–397.

Altuğ, Sumru and Pamela Labadie, *Dynamic Choice and Asset Markets*, Academic Press, 1994.

Anderson, T.W. and I. Olkin, "Maximum-likelihood Estimation of the Parameters of a Multivariate Normal Distribution," *Linear Algebra and its Applications*, 1985, *70*, 147 – 171.

Antonovics, Kate and Limor Golan, "Experimentation and Job Choice," *Journal of Labor Economics*, 2012, *30* (2), 333–366.

Arcidiacono, Peter and Robert A. Miller, "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 2011, *79* (6), 1823–1867.

_ and _ , "Identifying Dynamic Discrete Choice Models off Short Panels," Working Paper 2015.

_ , Esteban Aucejo, Arnaud Maurel, and Tyler Ransom, "College Attrition and the Dynamics of Information Revelation," Working Paper, NBER 2016.

Baumol, William J., "Entrepreneurship, Innovation and Growth: The David-Goliath Symbiosis," *Journal of Entrepreneurial Finance*, 2002, *7* (2).

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

Busenitz, Lowell W., "Entrepreneurial Risk and Strategic Decision Making: It?s a Matter of Perspective," *The Journal of Applied Behavioral Science*, 1999, *35* (3), 325–340.

Camerer, Colin and Dan Lovallo, "Overconfidence and Excess Entry: An Experimental Approach," *The American Economic Review*, 1999, *89* (1), 306–318.

Chatterji, Aaron, Edward L. Glaeser, and William R. Kerr, "Clusters of Entrepreneurship and Innovation," in Josh Lerner and Scott Stern, eds., *Innovation Policy and the Economy, National Bureau of Economic Research*, Vol. 14, University of Chicago Press, 2013, chapter 5, pp. 129–166.

DeGroot, Morris H., *Optimal Statistical Decisions*, McGraw Hill, 1970.

Dillon, Eleanor W. and Christopher T. Stanton, "Self-Employment Dynamics and the Returns to Entrepreneurship," Working Paper 2016.

Evans, David S. and Boyan Jovanovic, "An estimated model of entrepreneurial choice under liquidity constraints," *The Journal of Political Economy*, 1989, pp. 808–827.

Fairlie, Robert W. and Javier Miranda, "Taking the Leap: The Determinants of Entrepreneurs Hiring Their First Employee," *Journal of Economics & Management Strategy*, Forthcoming.

Gayle, George-Levi and Robert A. Miller, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?," *American Economic Review*, 2009, *99*, 1740–1769.

_ , Limor Golan, and Robert A. Miller, "Promotion Turnover and the Compensation in the Executive Labor Market," *Econometrica*, 2015, *83* (6), 2293–2369.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent, "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 2005, *23* (4), 681–724.

Gottlieb, Joshua D., Richard R. Townsend, and Ting Xu, "Experimenting with Entrepreneurship: The Effect of Job Protected Leave," Working Paper, NBER 2016.

Green, Edward J., "Lending and the Smoothing of Uninsurable Income," in E. Prescott and N. Wallace, eds., *Contractual Arrangements for Intertemporal Trade*, University of Minnesota Press, 1987, chapter 1, pp. 3–25.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Does Local Financial Development Matter," *The Quarterly Journal of Economics*, 2004, pp. 929–969.

Hamilton, Barton H., "Does entrepreneurship pay? An empirical analysis of the returns to self-employment," *The Journal of Political Economy*, 2000, *108* (3), 604–631.

— , **Nidhi Pande, and Nicholas W. Papageorge**, "The Right Stuff? Personality and Entrepreneurship," Mimeo, Johns Hopkins University 2016.

Heckman, James J. and Bo E. Honoré, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58* (5), 1121–1149.

Hotz, V. Joseph and Robert A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Humphries, John Eric, "The Causes and Consequences of Self-Employment over the Life Cycle," Mimeo, University of Chicago 2016.

Hurst, Erik and Annamaria Lusardi, "Liquidity constraints, household wealth, and entrepreneurship," *The Journal of Political Economy*, 2004, *112* (2), 319–347.

— **and Benjamin W. Pugsley**, "What Do Small Businesses Do?," *Brookings Papers on Economic Activity*, 2011, *2.*

— **and** — , "Wealth, Tastes, and Entrepreneurial Choice," Working Paper, NBER 2015.

Iyigun, Murat F. and Ann L. Owen, "Risk, Entrepreneurship, and Human-Capital Accumulation," *The American Economic Review*, 1998, *88* (2), 454–457.

James, Jonathan, "Ability Matching and Occupational Choice," SSRN Scholarly Paper ID 1937879, Social Science Research Network, Rochester, NY October 2011.

Jamshidian, Mortanza and Robert I. Jennrich, "Standard Errors for EM Estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2000, *62* (2), 257–270.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *The Journal of Political Economy*, 1979, *87* (5), 972–990.

— **and Yaw Nyarko**, "Learning by Doing and the Choice of Technology," *Econometrica*, 1996, *64* (6), 1299–1310.

Kawaguchi, Daiji, "Human Capital Accumulation of Salaried and Self-employed Workers," *Labour Economics*, 2003, *10* (1), 55–71.

Keane, Michael P. and Kenneth I. Wolpin, "The Career Decisions of Young Men," *The Journal of Political Economy*, 1997, *105* (3), 473–522.

Kreider, Rose M. and Renee Ellis, "Number, Timing, and Duration of Marriages and Divorces: 2009," Technical Report, U.S. Census Bureau. 2011.

Lafontaine, Francine and Kathryn Shaw, "Serial Entrepreneurship: Learning by Doing?," *Journal of Labor Economics*, 2016, *34* (S2), S217–S254.

Lazear, Edward P., "Entrepreneurship," *Journal of Labor Economics*, 2005, *23* (4), 649–680.

Levine, Ross and Yona Rubinstein, "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?," *Quarterly Journal of Economics*, Forthcoming.

Liang, James, Hui Wang, and Edward P. Lazear, "Demographics and Entrepreneurship," Working Paper, NBER 2014.

Magnac, Thierry and David Thesmar, "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 2002, *70* (2), 801–816.

Manso, Gustavo, "Experimentation and the Returns to Entrepreneurship," SSRN Scholarly Paper ID 2527034, Social Science Research Network, Rochester, NY November 2014.

March, James G. and Zur Shapira, "Managerial Perspectives on Risk and Risk Taking," *Management Science*, 1987, *33* (11), 418–438.

Margiotta, Mary M. and Robert A. Miller, "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 2000, *41* (3), 418–438.

Miller, Robert A., "Job Matching and Occupational Choice," *The Journal of Political Economy*, December 1984, *92* (6), 1086–1120.

Munk, Robert, "Does Self-Employment Pay? The Role of Timing and Voluntary Mobility," Mimeo, Ohio State University 2015.

Nanda, Ramana and Jesper B. Sørensen, "Workplace Peers and Entrepreneurship," *Management Science*, 2010, *56* (7), 1116–1126.

OECD, "Supporting Youth in Entrepreneurship," Seminar Summary, OECD, Paris 2013.

Oosterbeek, Hessel, Mirjam van Praag, and Auke Ijsselstein, "The Impact of Entrepreneurship Education on Entrepreneurship Skills and Motivation," *European Economic Review*, 2010, *54*, 442–454.

Peterman, Nicole E. and Jessica Kennedy, "Enterprise Education: Influencing Students? Perceptions of Entrepreneurship," *Entrepreneurship Theory and Practice*, 2003, *28* (2), 129–144.

Roll, Richard, "The Hubris Hypothesis of Corporate Takeovers," *The Journal of Business*, 1986, *59* (2), 1116–1126.

Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers (New Series)*, 1951, *3* (2), 135–146.

Rust, John, "Structural Estimation of Markov Decision Processes," in R. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: North Holland, 1994, chapter 51, pp. 3081–3143.

Ryan, Camille and Julie Siebens, "Educational Attainment in the United States: 2009," Technical Report, U.S. Census Bureau. 2012.

Sanders, Carl, "Skill Uncertainty, Skill Accumulation, and Occupational Choice," Mimeo, Washington University in St. Louis 2010.

Schumpeter, Joseph A., *The Theory of Economic Development*, Cambridge, MA: Harvard University Press, 1911.

Shane, Scott, "Why Encouraging More People to Become Entrepreneurs is Bad Public Policy," *Small Business Economics*, 2009, *33*, 141–149.

Souitaris, Vangelis, Stefania Zerbinati, and Andreas Al-Laham, "Do Entrepreneurship Programmes Raise Entrepreneurial Intention of Science and Engineering Students? The Effect of Learning, Inspiration and Resources," *Journal of Business Venturing*, 2007, *22*, 566–591.

von Graevenitz, Georg, Dietmar Harhoff, and Richard Weber, "The Effects of Entrepreneurship Education," *Journal of Economic Behavior & Organization*, 2010, *76*, 90–112.

8 Figures and Tables

Table 1: Summary Statistics: Individuals

Individuals	1506
Black	0.22
College or more	0.42
Age at entry	21.88
	[1.96]
Permanent Wealth	400
	[674]

Notes: Permanent wealth is in thousands of dollars of 2000. Standard deviation is in square brackets.

Table 2: Summary Statistics: Individual-Years

	All	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
Observations	21334	8902	9957	1403	602	470
%	100.00	41.73	46.67	6.58	2.82	2.20
Marital Status	0.76	0.74	0.77	0.79	0.86	0.50
High School	0.28	0.50	0.10	0.22	0.13	0.52
Some College	0.28	0.35	0.22	0.29	0.24	0.23
College	0.21	0.10	0.30	0.21	0.29	0.08
Some Grad	0.23	0.05	0.39	0.27	0.34	0.17
Age	31.04	28.92	32.21	33.93	36.94	30.45
	[7.27]	[6.65]	[7.18]	[7.35]	[7.00]	[8.06]
Wrkhrs	2147	2096	2234	2329	2703	
	[693]	[617]	[559]	[819]	[724]	
Hr Labor Income	18.71	14.16	21.24	21.30	37.91	
	[16.12]	[7.97]	[14.26]	[23.08]	[51.69]	
Residual						
Hr Labor Income		[6.99]	[12.32]	[20.41]	[44.29]	

Notes: White collar occupations are: professional, technical, and kindred workers; managers and administrators, except farm related; sales workers; clerical and kindred workers. Blue collar occupations are: craftsmen and kindred workers; operatives, except transport (including armed forces); transport equipment operatives; laborers, except farm related; service workers, except private household. Farm related occupations and military personnel dropped. Individuals are classified as unemployed if they reported to be not working or working for less than 2.5% the total amount of available hours in a year. Monetary quantities are in real dollars of 2000. Standard deviation is in square brackets. Residual income computed from occupation-specific OLS regressions on race, education and second degree polynomials on occupation-specific experience. One unit of hourly income represents 10$ per hour.

Table 3: Transition Matrix

	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
Blue Collar	0.87	0.09	0.02	0.00	0.02
White Collar	0.07	0.89	0.02	0.01	0.01
Unincorporated	0.10	0.10	0.74	0.04	0.01
Incorporated	0.03	0.14	0.07	0.76	0.01
Unemployed	0.37	0.16	0.04	0.00	0.43

Notes: Matrix entry i, j represents the proportion of people in occupation in row i who move into occupation in column j between t and $t + 1$.

Table 4: First Entry

	Blue Collar	White Collar	Unincorporated	Incorporated
Ever	0.65	0.87	0.28	0.15
At First Entry				
Age	23.16	25.60	32.23	35.50
exp_{bc}	-	2.81	3.88	2.42
exp_{wc}	1.30	-	5.13	8.44
exp_{eu}	0.11	0.14	-	1.38
exp_{ei}	0.02	0.04	0.52	-

Notes: Statistics computed using individuals that are observed from the beginning of their careers until at least age 40. This leaves 486 unique individuals. No observations are used beyond age 50.

Table 5: Occupation Spells

	All	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
Total	4294	1707	1652	453	194	288
Percent		39.75	38.47	10.55	4.52	6.71
Duration	4.97	5.21	6.03	3.10	3.10	1.63
First		52.06	42.56	2.19	0.27	2.92

Notes: **Duration** is the average duration of spells in years. **First** is the percentage of first spells that belong to a particular occupation. **Tried** is the percentage of individuals who tried the occupation during their observed careers.

Table 6: Income Parameters

	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Constant	0.927**	0.008	0.847**	0.015	1.304**	0.059	0.730**	0.244
Black	-0.204**	0.008	-0.122**	0.014	-0.180**	0.052	-0.097	0.278
Some College	0.222**	0.009	0.129**	0.014	-0.095**	0.041	0.597**	0.237
College	0.323**	0.017	0.456**	0.018	0.537**	0.083	0.411*	0.216
More than College	0.251**	0.021	0.624**	0.020	0.934**	0.080	1.842**	0.366
Married	0.053**	0.005	0.213**	0.008	0.019	0.039	0.718**	0.114
$1\{exp_{bc} \geq 1\}$	0.081**	0.006	-0.077**	0.011	-0.233**	0.053		
$1\{exp_{bc} \geq 2\}$	0.074**	0.007					-0.268	0.179
$1\{exp_{bc} \geq 3\}$	0.041**	0.007						
$1\{exp_{bc} \geq 4\}$	0.056**	0.008						
$1\{exp_{bc} \geq 5\}$	0.067**	0.008					-0.508	0.319
$1\{exp_{bc} \geq 6\}$	0.142**	0.007						
$1\{exp_{bc} \geq 7\}$			0.119**	0.015				
$1\{exp_{wc} \geq 1\}$	0.133**	0.006	0.126**	0.012				
$1\{exp_{wc} \geq 2\}$			0.127**	0.013	0.019	0.063	-0.765**	0.182
$1\{exp_{wc} \geq 3\}$			0.192**	0.012	0.044	0.072	0.824**	0.228
$1\{exp_{wc} \geq 5\}$			0.140**	0.011				
$1\{exp_{wc} \geq 6\}$	0.209**	0.013			0.480**	0.061		
$1\{exp_{wc} \geq 7\}$			0.178**	0.014				
$1\{exp_{wc} \geq 8\}$			0.410**	0.013			0.627**	0.124
$1\{exp_{eu} \geq 1\}$			-0.177**	0.016	0.219**	0.033		
$1\{exp_{eu} \geq 3\}$							0.229	0.241
$1\{exp_{eu} \geq 4\}$			0.743**	0.032	0.407**	0.034		
$1\{exp_{eu} \geq 7\}$					0.280**	0.038		
$1\{exp_{ei} \geq 1\}$					-0.451**	0.060	0.433**	0.086
$1\{exp_{ei} \geq 2\}$			1.004**	0.028				
$1\{exp_{ei} \geq 3\}$					1.867**	0.109		
$1\{exp_{ei} \geq 5\}$							1.078**	0.085
$1\{exp_{e} \geq 1\}$	-0.022*	0.011						
$1\{exp_{e} \geq 2\}$	-0.156**	0.017						
$1\{exp_{e} \geq 5\}$	0.365**	0.026						
Obs	8902		9957		1403		602	

Notes: Hourly income measured in $10s. *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. Estimated parameters of equation (4). Returns to experience are estimated as step functions. As an example, $1\{exp_{eu} \geq 3\}$ indicates that the individual has three years or more of unincorporated experience. In blue collar work, experience from both entrepreneurial occupations is pooled: $exp_e = exp_{eu} + exp_{ei}$. Steps functions were chosen to avoid out of sample return estimates especially on entrepreneurial occupations. The steps were chosen using statistical significance in a preliminary OLS.

Table 7: Population Ability Covariance Matrices

High School

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.15**	0.003						
white collar	0.13**	0.007	0.11**	0.008				
unincorporated	0.20**	0.020	0.17**	0.029	0.27**	0.040		
incorporated	0.04	0.479	0.03	0.313	0.05	0.920	0.04	0.293

Some College

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.23**	0.006						
white collar	0.26**	0.009	0.32**	0.009				
unincorporated	0.16**	0.032	0.14**	0.032	0.28**	0.033		
incorporated	0.45**	0.128	0.77**	0.071	0.35*	0.179	4.60**	0.395

College

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.45**	0.021						
white collar	0.33**	0.026	0.57**	0.016				
unincorporated	0.29**	0.099	0.38**	0.059	3.52**	0.249		
incorporated	0.71**	0.079	0.85**	0.038	-0.11	0.528	1.66**	0.179

More than College

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.37**	0.020						
white collar	0.22**	0.035	0.87**	0.022				
unincorporated	-0.41**	0.110	0.66**	0.075	3.03**	0.175		
incorporated	-0.29	0.685	1.82**	0.149	2.35**	0.706	10.88**	0.665

Notes: *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. Covariance matrix of the joint distribution of unobserved ability conditional on education, denoted $\boldsymbol{\Delta}_s$.

Table 8: Idiosyncratic Variance

blue collar		white collar		unincorporated		incorporated	
coeff	se	coeff	se	coeff	se	coeff	se
0.30**	0.001	0.96**	0.004	2.47**	0.03	8.00**	0.134

Notes: *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. Idiosyncratic hourly income variance in every occupation, denoted $\sigma^2_{\eta_k}$.

Table 9: Utility Parameters

ρ	coeff	se
	0.040**	0.0005

High School		blue collar		white collar		unincorporated		incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-1.804**	0.042	-0.939**	0.048	-0.549**	0.061	-1.459**	0.104
	black	0.738**	0.041	1.173**	0.047	0.895**	0.061	1.724**	0.135
	married	-0.684**	0.040	-0.364**	0.047	-0.544**	0.060	-0.051	0.112
Entry Cost	*constant*	-4.723**	0.338	2.975**	0.143	5.298**	0.154	12.645**	0.251
	$age/10$	2.837**	0.151	-0.132*	0.053	-0.536**	0.054	-2.090**	0.084
	$\omega_i/10^3$	2.681**	0.705	-1.529**	0.377	0.465*	0.231	-0.258	0.196
	$(age/10)\cdot(\omega_i/10^3)$	-1.108**	0.287	0.421**	0.143	-0.234**	0.080	0.083	0.067

Some College		blue collar		white collar		unincorporated		incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-1.986**	0.052	-1.920**	0.053	-1.717**	0.066	-1.361**	0.090
	black	0.466**	0.058	0.857**	0.061	1.069**	0.081	1.248**	0.109
	married	-0.933**	0.058	-0.784**	0.060	-0.686**	0.071	-0.874**	0.099
Entry Cost	*constant*	-3.838**	0.213	3.807**	0.171	4.974**	0.148	11.559**	0.200
	$age/10$	2.061**	0.087	-0.109	0.067	-0.331**	0.050	-1.649**	0.060
	$\omega_i/10^3$	-0.515	0.413	-2.989**	0.456	0.024	0.077	-0.526**	0.069
	$(age/10)\cdot(\omega_i/10^3)$	0.236	0.168	1.085**	0.188	-0.029	0.023	0.158**	0.017

College		blue collar		white collar		unincorporated		incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-2.667**	0.101	-3.596**	0.100	-2.987**	0.113	-4.394**	0.142
	black	0.216*	0.104	1.465**	0.104	1.755**	0.169	16.547	319.588
	married	-0.208	0.107	0.381**	0.106	0.067	0.119	1.002**	0.146
Entry Cost	*constant*	-3.117**	0.390	1.199**	0.291	5.810**	0.215	7.482**	0.288
	$age/10$	1.887**	0.149	0.771**	0.110	-0.534**	0.068	-0.700**	0.085
	$\omega_i/10^3$	-1.243	0.957	-1.912**	0.729	-3.583**	0.341	-2.451**	0.297
	$(age/10)\cdot(\omega_i/10^3)$	0.800*	0.372	0.375	0.287	1.106**	0.127	0.610**	0.102

More than College		blue collar		white collar		unincorporated		incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-1.038**	0.066	-1.902**	0.063	-1.287**	0.077	-0.457**	0.109
	black	-0.637**	0.116	-0.298**	0.114	0.423*	0.166	-0.809**	0.144
	married	-0.436**	0.076	-0.336**	0.072	-0.740**	0.085	-0.966**	0.117
Entry Cost	*constant*	-5.888**	0.442	-3.414**	0.479	5.186**	0.238	9.253**	0.301
	$age/10$	2.689**	0.166	2.481**	0.187	-0.200**	0.072	-0.875**	0.086
	$\omega_i/10^3$	1.951*	0.809	1.712	0.897	-0.531	0.284	-3.499**	0.241
	$(age/10)\cdot(\omega_i/10^3)$	-0.459	0.301	-0.921**	0.357	0.070	0.088	0.897**	0.074

Notes: *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. ω_i is defined as the individual's permanent wealth in Section 2 and it is measured in thousands of dollars of 2000. Estimated parameters of equations (2) and (3).

Table 10: First Entry: Observed and Simulated

Data

	blue collar	white collar	unincorporated	incorporated
Tried by age 40	0.65	0.83	0.23	0.11
At first entry				
Age	22.84	24.81	29.57	32.82
exp_{bc}	-	1.99	3.48	2.05
exp_{wc}	1.08	-	3.37	6.58
exp_{eu}	0.03	0.15	-	0.96
exp_{ei}	0.01	0.02	0.22	-

Model

	blue collar	white collar	unincorporated	incorporated
Tried by age 40	0.77	0.73	0.23	0.09
At first entry				
Age	22.23	25.84	29.98	32.40
exp_{bc}	-	3.32	4.51	3.73
exp_{wc}	0.58	-	3.17	5.53
exp_{eu}	0.08	0.25	-	0.48
exp_{ei}	0.00	0.04	0.11	-

Notes: Statistics computed using individuals that are observed from the beginning of their careers until at least age 40. Only data when individuals are 40 years old or below are used.

Table 11: Young Incorporated Entrepreneurship Subsidy

	Subsidy in $1000s		
	0	25	50
Young Entrepreneurs			
Tried in first 5 years	0.02	0.03	0.05
Mean belief ($ per hour) at 1st entry	3.7	2.5	1.2
Mean ability ($ per hour) at 1st entry	5.0	3.6	2.1
Bias (belief-abiliity)	-1.3	-1.1	-0.9
95th pctile ability ($ per hour) at 1st entry	46.5	44.0	37.8
All entrepreneurs			
Tried	0.15	0.17	0.22
Participation rate at age 40	0.04	0.05	0.06
PVI net of subsidy ($1000s)	757	770	770
Mean belief ($ per hour) at 1st entry	6.4	6.2	5.2
Mean ability ($ per hour) at 1st entry	5.5	5.4	4.6
Bias (belief-abiliity)	0.9	0.8	0.6
All individuals			
PVI net of subsidy ($1000s)	508	513	526

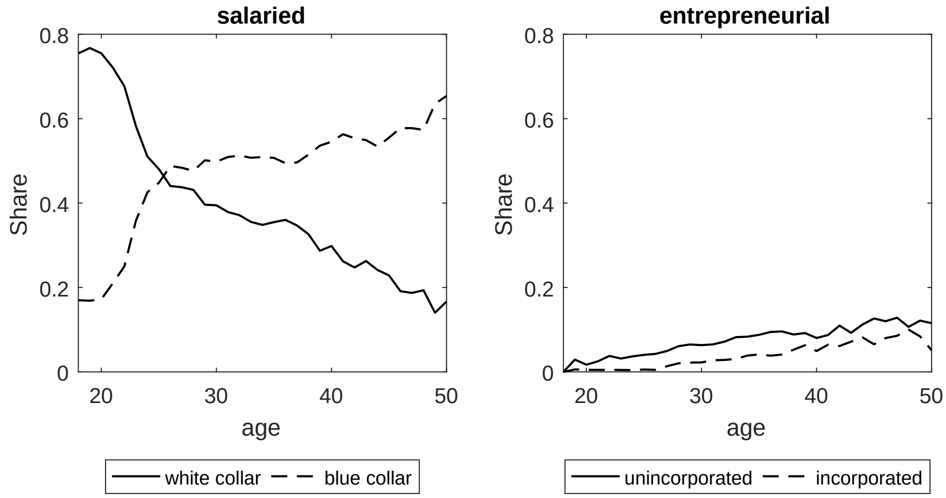
Notes: Subsidy given only to individuals who decide to attempt incorporated entrepreneurship immediately after finishing their education. Young entrepreneurs are those who tried incorporated entrepreneurship for the first time within their first five years in the labor market. *Rows*: Summary statistics are computed separately for young incorporated entrepreneurs and for all incorporated entrepreneurs. PVI stands for the present value of income. *Columns*: The no-subsidy column corresponds to the baseline model.

Table 12: Young Incorporated Entrepreneurship Education

	Noise Variance Scale, s				
	inf	10	5	2	1
Young Entrepreneurs					
Tried in first 5 years	0.02	0.12	0.11	0.10	0.08
Mean belief ($ per hour) at 1st entry	3.7	82.4	75.1	61.3	48.4
Mean ability ($ per hour) at 1st entry	5.0	1.3	2.1	5.5	10.7
Bias (belief-abiliity)	-1.3	81.1	73.0	55.8	37.7
95th pctile ability ($ per hour) at 1st entry	46.5	47.0	47.5	49.9	59.1
Overall					
Tried	0.15	0.26	0.26	0.25	0.24
Participation rate at age 40	0.04	0.12	0.12	0.12	0.11
PVI ($1000s)	757	944	941	961	983
Mean belief ($ per hour) at 1st entry	6.4	45.9	41.8	33.8	26.5
Mean ability ($ per hour) at 1st entry	5.5	2.0	2.3	4.0	6.6
Bias (belief-abiliity)	0.9	43.9	39.5	29.8	19.9
All individuals					
PVI ($1000s)	508	573	575	578	581

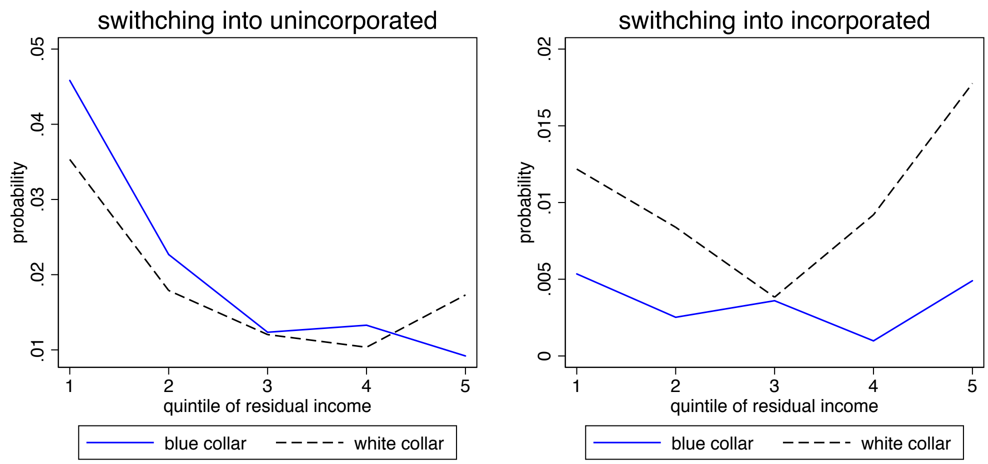
Notes: Individual-specific signal about incorporated ability given to everybody immediately after finishing their education. Interventions are characterized by the noise variance of their signals (σ_ν), expressed in terms of the noise variance of trying incorporated entrepreneurship in reality (σ_{η_4} in Table 8): $\sigma_\nu = s \cdot \sigma_{\eta_4}$. Young entrepreneurs are those who tried incorporated entrepreneurship for the first time within their first five years in the labor market. *Rows*: Summary statistics are computed separately for young incorporated entrepreneurs and for all incorporated entrepreneurs. PVI stands for the present value of income. *Columns*: The *inf* column corresponds to the baseline model where no signal is received. This column is identical to the no-subsidy column in Table 11.

FIGURE 1: Occupational Choice: Age Profile



Notes: Participation rates into each occupation by age.

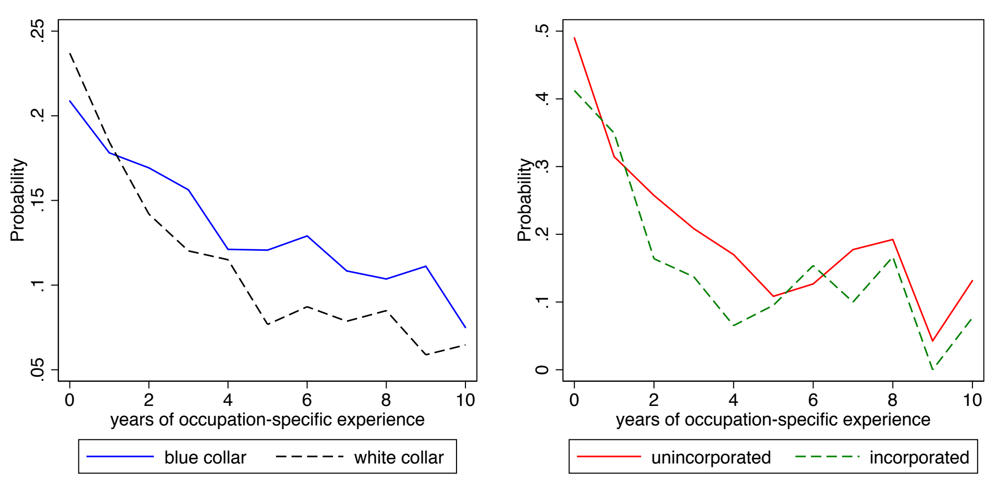
Figure 2: Probability of Switching into Entrepreneurial Occupations



Notes: Probability of switching into entrepreneurial occupations in $t+1$ by decile of residual income in t. Residual income is computed from occupation-specific regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status.

.

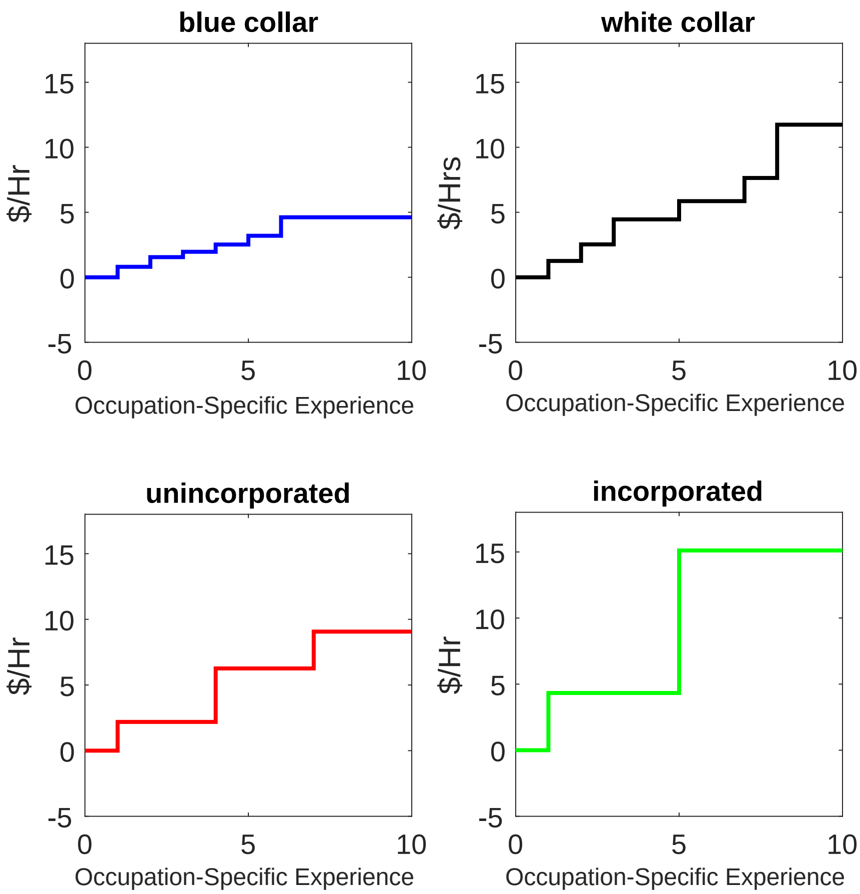
Figure 3: Probability of Switching by Occupation-Specific Experience



Notes: Probability of switching from occupation k next period conditional on years of experience in occupation k. Figure considers only individuals who are observed for at least 10 years in the sample.

.

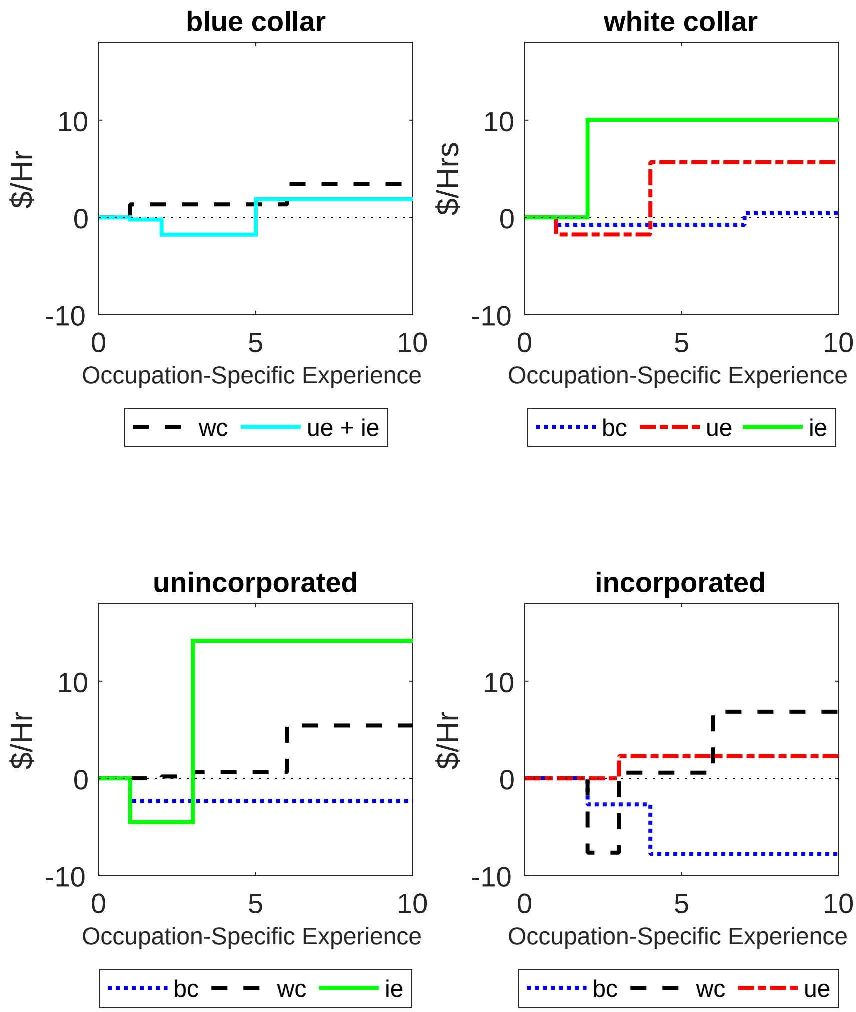
Figure 4: Returns to Own-Occupation Experience



Notes: Returns implied by the marginal step estimates in Table 6.

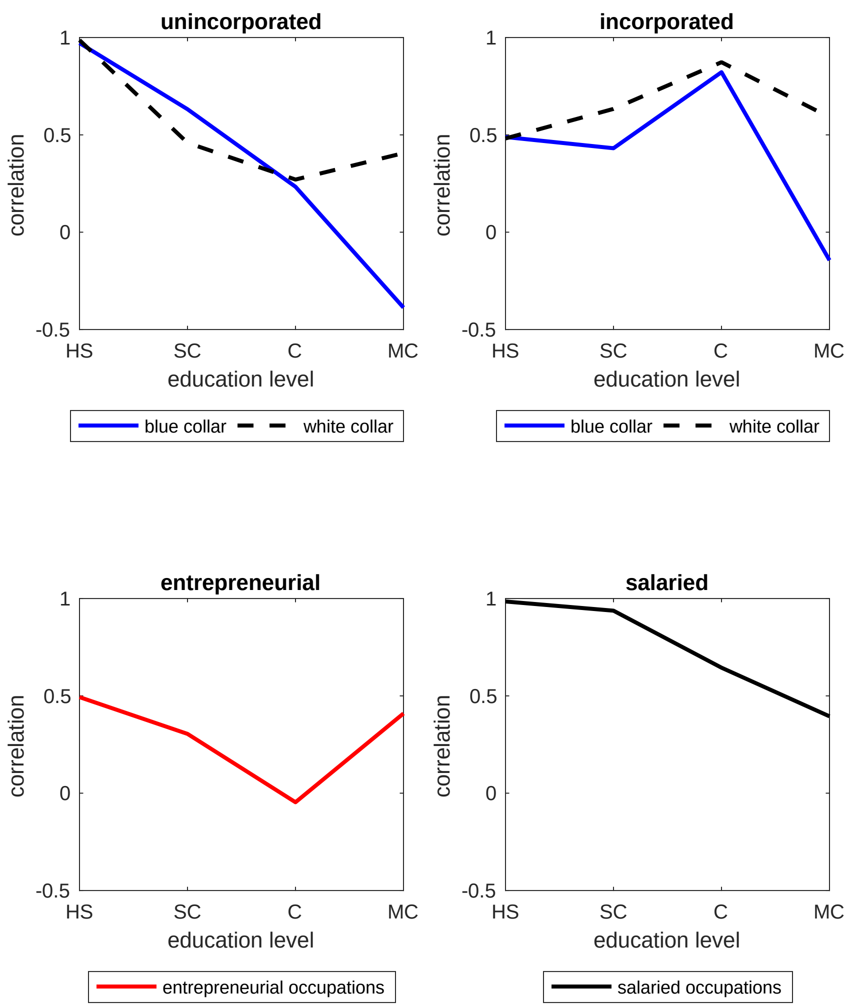
.

FIGURE 5: Returns to Cross-Occupation Experience



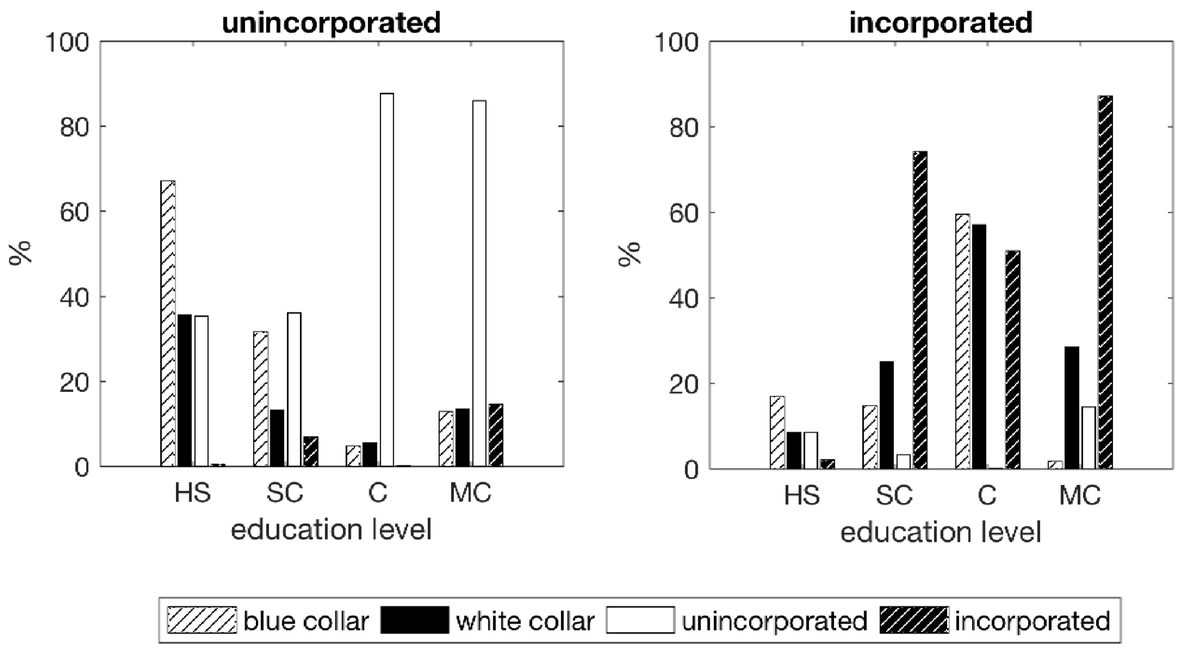
Notes: Returns implied by the marginal step estimates in Table 6. Occupations are: white collar (wc), blue collar (bc), unincorporated entrepreneurship (ue), and incorporated entrepreneurship (ie). In blue collar work, experience from both entrepreneurial occupations is pooled.

Figure 6: Implied Correlation Between Abilities



Notes: Correlation between abilities per education level implied by the estimates in Table 7. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). Top figures show the implied correlations of unobserved ability between salaried occupations and each of the entrepreneurial occupations. Bottom left figure shows the implied correlation of abilities between unincorporated and incorporated entrepreneurship. Bottom right figure shows the implied correlation of abilities between blue and white collar.

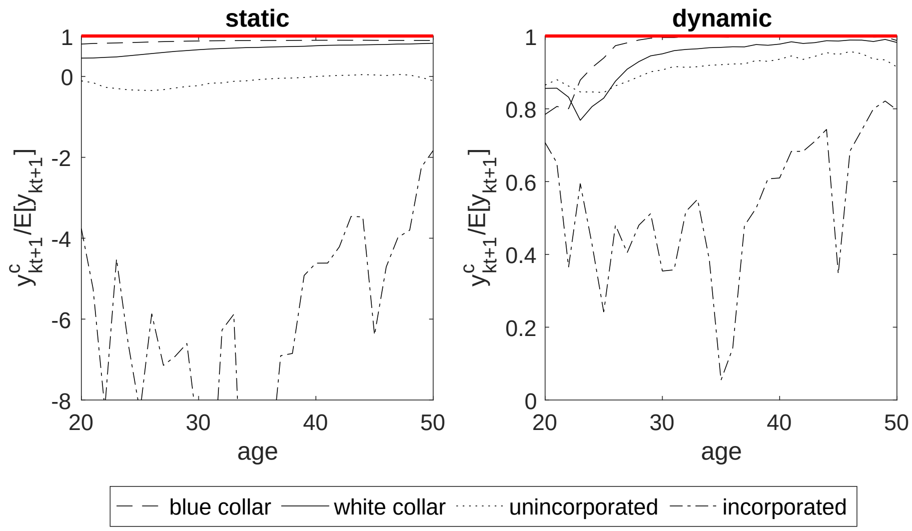
FIGURE 7: Prior Variance Eliminated After 5 Years of Occupation-Specific Experience



Notes: Figure shows how much prior variance remains in occupation k after accumulating five years of experience in occupation k' and zero years in any other occupation. On the y-axis of the panel devoted to occupation k is the percent quantity $(1 - \mathbb{V}_{\{k,k\},5}(k')/\mathbb{V}_{\{k,k\},0}) * 100$. The numerator is the belief variance of occupation k after accumulating five years of experience in occupation k' and zero in any other occupation. The denominator is the prior variance of occupation k. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). Quantities are obtained using the estimates in Table 7.

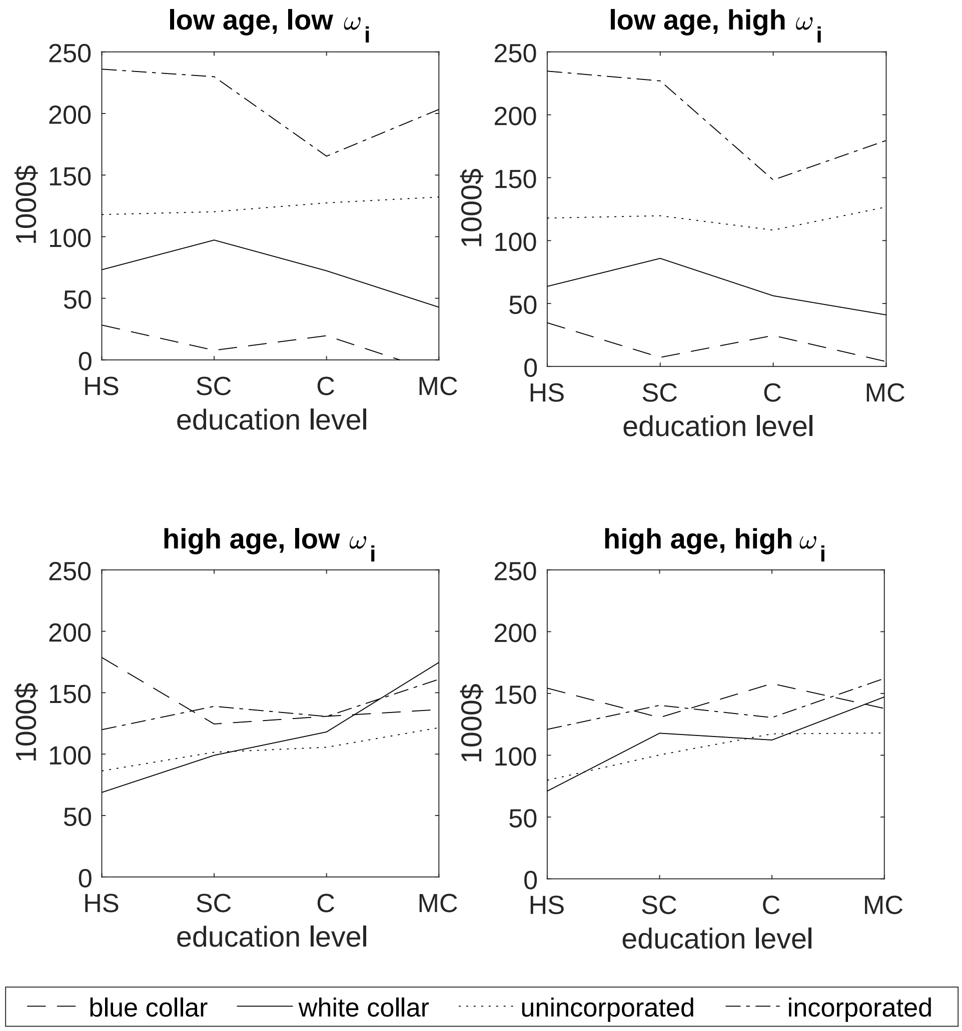
.

Figure 8: Certainty Equivalent



Notes: On the y-axis of both figures is the average of a scale free measure of risk aversion computed across individuals of a given age with positive expected income, $E[y_{kt+1}|h_t, \mathbb{E}_t] > 0$. The measure can be understood as the proportion of expected annual income that would be necessary to let a static/dynamic individual indifferent between taking the gamble of going into the occupation and receiving the certainty equivalent for sure without entering the occupation. In the numerator is the certainty equivalent, y^c_{kt+1}, static and dynamic, obtained in equations (48) and (49) in the Results appendix. In the denominator is the expected hourly income conditional on beliefs, $E[y_{kt+1}|h_t, \mathbb{E}_t]$.

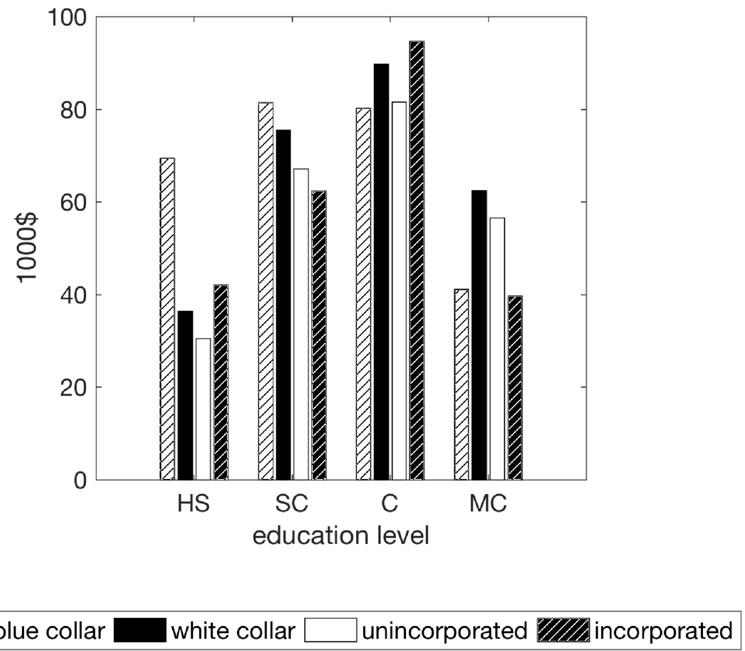
FIGURE 9: Monetary Equivalent of Entry Costs



Notes: On the x-axis of each panel is the education level. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). On the y-axis is the monetary equivalent of entry costs, obtained using the estimates in Table 9 and equation (50) in the Results appendix. The top two panels correspond to individuals 20 years old, while the bottom two correspond to individuals 40 years old. The left panels correspond to individuals in the 10th percentile of permanent wealth ω_i, while the right panels correspond to individuals in the 90th percentile.

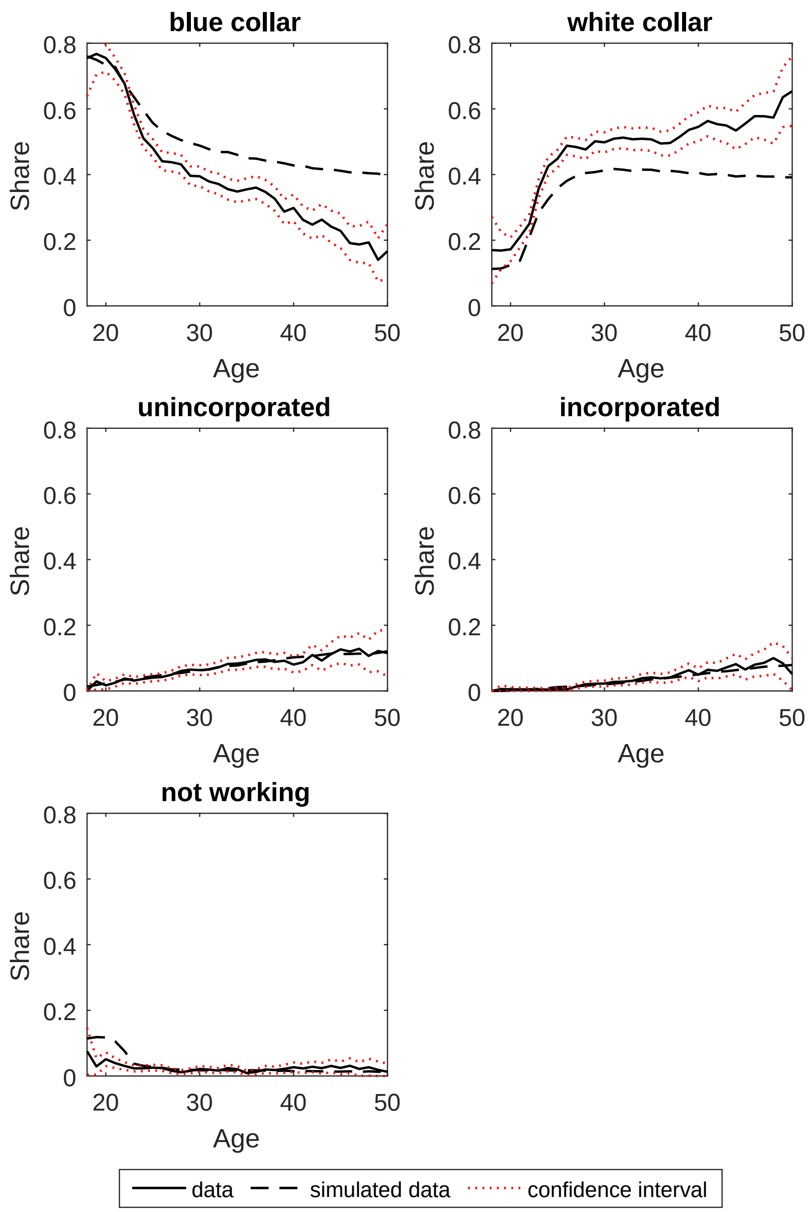
.

Figure 10: Monetary Equivalent of non-pecuniary Benefits



Notes: On the x-axis are levels education: high school (HS), some college (SC), college (C), and more than college (MC). Occupations are: white collar (wc), blue collar (bc), unincorporated entrepreneurship (ue), and incorporated entrepreneurship (ie). On the y-axis is the monetary equivalent of the non-pecuniary benefits not related to entry. They are obtained using the estimates in Table 9 and a similar derivation as in equation (50) in the Results appendix. The benefits are computed for a white, married man.

Figure 11: Simulated versus Observed Choice Rates



Notes: Actual and simulated choices by age.

FIGURE 12: Decomposition: Entry

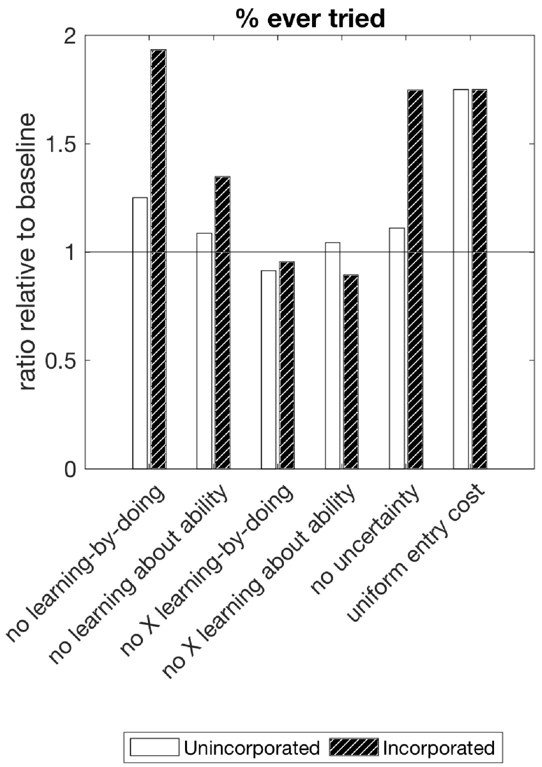


FIGURE 13: Decomposition: Timing

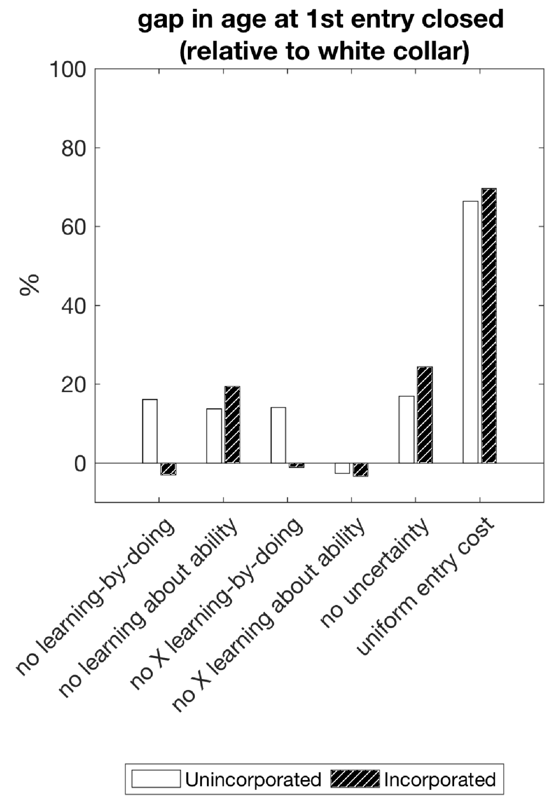


FIGURE 14: Decomposition: Ability

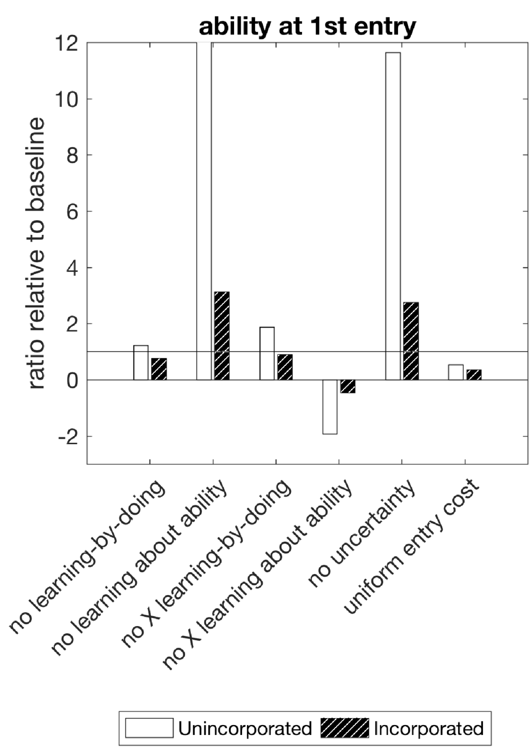
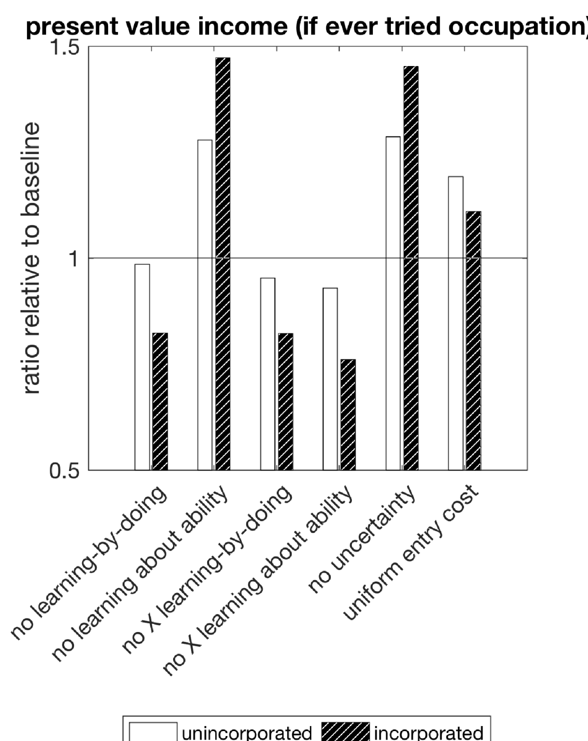


FIGURE 15: Decomposition: PVI



A Data Appendix

A.1 PSID Data

This paper uses data from the Panel Study of Income Dynamics (PSID). The PSID started in 1968 with a representative sample of about 18,000 individuals in 5,000 families in the United States. Information about these individuals and their descendants was collected yearly up to 1996, year after which the study became biennial. The study is restricted to white and black men between years 1968 and 1996. Survey information used include data on occupation, self-employment status, business ownership, incorporation status, labor income, business income, working hours, completed education, age, race, and marital status.

Setting the beginning of labor market careers. In order to account for the process of belief formation individuals must be observed from their entrance into the labor market. Potential experience is defined as

$$PotentialExperience = Age - CompletedEducation - 6$$

to set the beginning of individuals' labor market careers. First, the minimum potential experience for each individual is computed. Only those individuals whose minimum potential experience is at most 3 are kept. Then, the beginning of the individual's labor market career is set whenever

$$PotentialExperience = \begin{cases} 0 & if \quad \min PotentialExperience \leq 0 \\ k & if \quad \min PotentialExperience = k \in \{1, 2, 3\} \end{cases}$$

Self-employment. At any period, conditional on having declared to be working (or working for money) or only temporarily laid off, individuals answer the following question (or a slightly modified version of it):

"*On your main job, are you self-employed, are you employed by someone else, or what?*"

The answer alternatives are "Someone else," "Both someone else and self," "Self-employed only," and "Don't Know." Entrepreneurs as defined as those individuals who have positive working hours and declare to be self-employed only. All other individuals with positive working hours are catalogued into one of the salaried occupations.

Occupation. The PSID provides the 3-digit occupation code from 1970 Census of Population which is build using the Alphabetical Index of Industries and Occupations issued June 1971 by the U.S. Department of Commerce and the Bureau of the Census was used for this variable. The PSID provides the following categorization of occupations

- ↪ Occupation 1: 1 - 195 Professional, Technical, and Kindred Workers
- ↪ Occupation 2: 201 - 245 Managers and Administrators, Except Farm
- ↪ Occupation 3: 260 - 285 Sales Workers
- ↪ Occupation 4: 301 - 395 Clerical and Kindred Workers
- ↪ Occupation 5: 401 - 600 Craftsmen and Kindred Workers
- ↪ Occupation 6: 601 - 695 Operatives, Except Transport
- ↪ Occupation 7: 701 - 715 Transport Equipment Operatives
- ↪ Occupation 8: 740 - 785 Laborers, Except Farm
- ↪ Occupation 9: 801 - 802 Farmers and Farm Managers
- ↪ Occupation 10: 821 - 824 Farm Laborers and Farm Foremen
- ↪ Occupation 11: 901 - 965 Service Workers, Except Private Household
- ↪ Occupation 12: 980 - 984 Private Household Workers

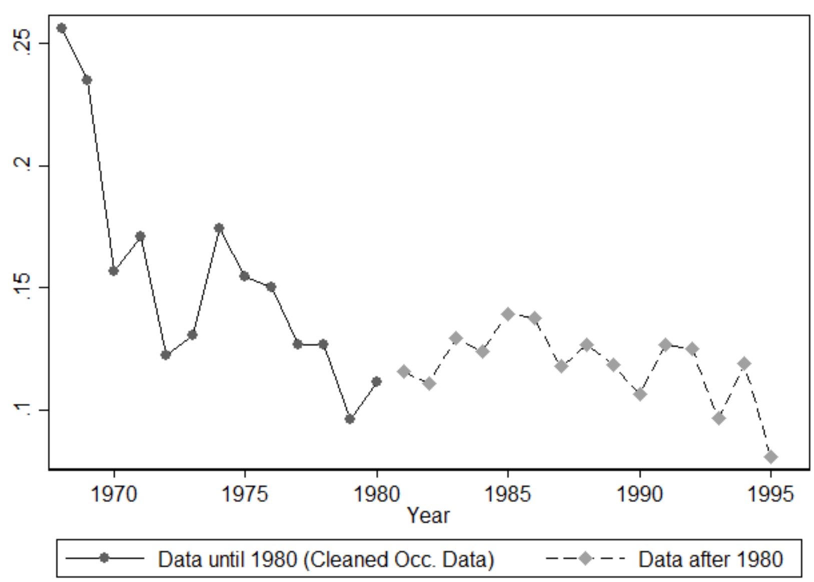
Observations corresponding to "members of the armed forces" (coded as 600) are dropped as well as observations of farm related occupations and observations of private household workers (occupations 9, 10 and 12). The remaining PSID categories are grouped into

- ↪ Blue Collar: Craftsmen and Kindred Workers; Operatives, Except Transport; Transport Equipment Operatives; Laborers, Except Farm; Service Workers, Except Private Household.
- ↪ White Collar: Professional, Technical, and Kindred Workers; Managers and Administrators, Except Farm; Sales Workers; Clerical and Kindred Workers

Individuals provide their occupation regardless of their self-employment status. However, provided that in the model entrepreneurship is an occupation on its own, the occupation data is disregarded whenever an individual is self-employed. Also, occupations were the individual reports working for more than 2.5% of the total amount of available hours in the year ($365.25 * 24$).

Up to 1980, the occupational data provided by the PSID is coded retroactively in order to correct for spurious transitions. PSID officials use original PSID reports and the three-digit 1970 Census occupation codes for a selected sample of PSID heads and spouses. Therefore, only part of the individuals' careers in the sample have been further corrected for spurious transitions. To the extent that the categories used in the paper are broad enough and that survey officials get more accurate cataloguing occupations over time, this problem should be minor in the sample. Figure 16 shows the switching trends among collar occupations, entrepreneurship, and unemployment, computed with data before and after 1980; the decrease in switching corresponds to the fact that individuals in the sample are acquiring more experience. More importantly, there seems to be no evidence of jumps in the trend of switching before and after at this level of aggregation.

FIGURE 16: Proportion of Individuals Switching between t and $t+1$ Over Years in the Labor Market (Collar)



Labor Income. The PSID labor income variable is computed equally for employed and self-employed individuals. Up to 1993, it corresponds in general to the sum of wages (before taxes or other deductions) and "actual amounts of labor part of farm income and business income, bonuses, overtime, commissions, professional practice, labor part of income from roomers and boarders, and market gardening" (PSID Codebook). From this variable the following components are subtracted: the labor part of business income, of farm income, and of income from roomers and boarders when available. Starting from 1994, the labor part of farm income and that of business income are not included in the variable. Labor income is bracketed for 1968 and 1969. The midpoint value of the bracket is assigned; however, less than 1% of the individual-year observations correspond to those years. Also, the PSID labor income variable is censored at different upper values at different years. Less than 0.2% percent of the observations correspond to censored observations. The labor part of farm income is bracketed until 1975. Again, the value of the midpoint of the bracket is assigned.

Business income. Business income is gathered for those individuals that satisfy the following two conditions:

↪ 1. They answer "yes" to the following question (or a modified version of it)

"*Did you or any other member of your family own a business at any time in year YYYY, or have a financial interest in any business enterprise?*"

Not all self-employed individuals answer "yes" to this question and not all individuals who answer yes to this question are self-employed. While about 82 percent of self-employed answer "yes" to this question less than 8 percent of salaried workers do. Regardless of this numbers, this may still be a drawback of how the paper treats the data.

↪ 2. They then proceed to say that the business mentioned was not uniquely a corporation. In other words, they proceed to say that the business was either (i) unincorporated or (ii) they have an interest in both types or (iii) they do not know.

If those two conditions are satisfied they then answer the question

"*How much was (your/his/her/their share of the total income from business in YYYY– that is, the amount (you/he/she/they) took out plus profit left in? [If zero: did you have a loss? How much was it?*"

Business income is computed as the sum of the labor and asset part of head's business income as reported in the PSID data. The labor part and asset part of business income are bracketed until 1975. Again, the value of the midpoint of the bracket is assigned. After computed, business income is added to the labor income measure only for unincorporated self-employed individuals.

Income. In summary, for salaried workers and incorporated self-employed individuals:

$$Income = LaborIncome$$

For self-employed unincorporated individuals:

$$Income = LaborIncome + BusinessIncome$$

Individuals who are not working any hours are assigned zero income. All values are in constant dollars of 2000.

Incorporated and unincorporated status. Following an affirmative answer to the business ownership question (above), individuals are asked about their incorporation status in all years in the PSID (denote this question IQ1). Also, in years 1975, 1976 and from 1985 onward, individuals are asked about their incorporated status after the self-employment question (denote this question IQ2). Even though question IQ2 seems closer to the paper's definition of entrepreneurship, not all years are available for this question. An imputation algorithm is followed in order to determine the incorporated status of entrepreneurs.

In the imputation algorithm more relevance is given to stability and consistency of the measure across years. The imputation entails the following steps: **(i)** Initially, the incorporation status of entrepreneurs is determined from question IQ1; **(ii)** If incorporated status for entrepreneurs is missing or ambiguous (individual reported "Both," "Other," or "Do not know") in IQ1, the value from question IQ2 is assigned insofar as it corresponds to "Incorporated" or "Unincorporated;" **(iii)** If data is still missing or ambiguous, the $t-5,\ldots,t-1,t+1,\ldots,t+5$ (past and future) answers from IQ1 and IQ2 are used to assign the incorporated status at t; **(iv)** all remaining ambiguous observations are imputed as "unincorporated." Out of 2201 observations of entrepreneurs, this method imputes 551 observations: 406 from step **(ii)**, 120 from step **(iii)** and 25 from step **(iv)**.

Working Hours. The study uses the individual level PSID variable for working hours. It counts the actual hours worked by the individual during year YYYY. Missing data were not assigned.

Hourly income. Hourly income is computed simply as annual income divided by annual working hours.

Education. Consistent with the procedure for setting the beginning of individuals' labor market careers, the education variable corresponds to the value of completed education. Education data are discretized into: high school (12 years of education or less), some college (13 to 15 years of education), college (16 years of education), more than college (more than 16 years of education). Education is censored at 17 years which may affect the potential experience criteria used above for setting the beginning of individuals' labor market careers.

Age and marital status. Reported age and marital status of individual.

Experience variables. Experience variables are computed using occupation data over the individual's career.

Wealth. The PSID includes a measure of wealth for selected years: 1984, 1989, 1994, and every two years starting in 1999. The wealth measure in the PSID is constructed as the sum of six types of assets (farm business, checking or savings accounts, real estate other than main home, stocks, vehicles, and other assets) net of debt value plus the value of home equity. Since the survey does not include data on wealth at every period, in the current analysis a measure of permanent wealth will be considered instead. In order to obtain the individual measure, the following fixed effects regression is run:

$$Wealth_{it} = \gamma_0 + \gamma_1 age_{it} + \gamma_2 age_{it}^2 + u_i + \epsilon_{it}$$

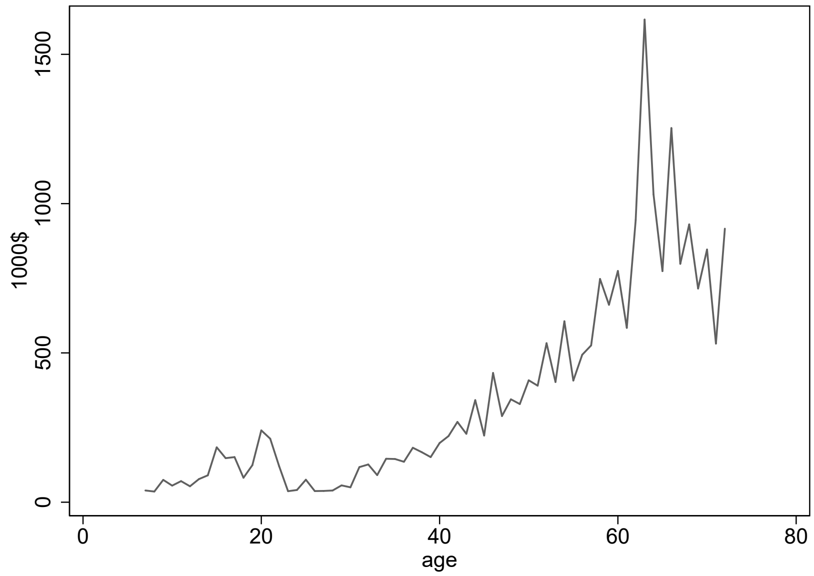
The individual measure for permanent wealth is then obtained as

$$\omega_i = \hat{\gamma}_0 + \hat{u}_i$$

In estimation only individuals with at least three wealth data points are used. Figure 17 shows the age profile of wealth accumulation from pooling all available data.

Full time vs part time workers. There is no differentiation in the treatment of the data between full-time workers and part time workers. In fact, only about 6% of individual-year observations for working individuals are part-time observations (less than 20 hours per week).

FIGURE 17: Average Wealth



Notes: Average wealth in thousands of dollars of 2000.

TABLE 13: Parameters of the Wealth Profile Equation

	coeff	se
γ_0	417.78	85.16
γ_1	-23.95	4.40
γ_2	0.47	0.06

Notes: wealth in thousands of dollars of 2000.

Part-time individual-year observations are not dropped because they would create gaps in the careers of 36% of the individuals (See Table 14 bellow).

TABLE 14: Full time and Part Time observations (1968-1996)

	Full time	Part Time
% of All individual-year worker observations	0.95	0.05
% of unique individuals ever in part time		0.35

Notes: Table shows proportion of individual-year observations with less than 1044 working hours in a year (part time) and proportion of unique individuals that were ever part time workers by this criterion. Selection rule was defined as not dropping part time observations since they would create gaps by about 40 percent of all unique individuals' labor histories.

Data Gaps. the histories of individuals with data gaps of more than 2 years are dropped.

For those with data in t and $t+2$ but not in $t+1$, Time is then redefined by making $t+1 = t+2$ and so forth. Similarly, for those with data in t and $t+3$ but not in $t+1$ and $t+2$, time is redefined by making $t+1 = t+3$ and so forth.

Dropping Data Process. Initial number of individuals: 75,260. Individuals remaining after dropping individuals with no information on age, 75,153 for 3'457,038 individual-year observations. Individual-year observations remaining after keeping only household heads and their spouses: 446,242;[65] individual-year observations remaining after keeping black or white individuals: 424,497; individual-year observations remaining after dropping years after 1996: 326,455; individual-year observations remaining after dropping females: 146,083; individual-year observations remaining after dropping missing participation info: 132,248; individual-year observations after dropping missing marital status: 132,242. Individual-year observations satisfying potential experience criterion: 37,759; individual-year observations remaining after dropping data on missing occupations, farm related occupations, and private household workers: 30,006; individual-year observations remaining after dropping missing income: 29,676; individual-year observations remaining after dropping military occupations: 28,683; individual-year observations remaining after dropping jumps in data: 26,087; individual-year observations satisfying potential experience criterion after previous droppings: 25,152; individual-year observations of people who never worked: 47. After dropping observations of individuals who lack data on relevant variables except wealth, data set contains 2,057 individuals and 25,105 individual-year observations. With this data set the first stage of the estimation procedure is undertaken. For the second stage, an extra dropping criterion is added to exclude those individuals with less than three data points of wealth. The final data set for estimation of the second stage contains 1,506 individuals and 21,334 individual-year observations.

A.2 Bond Price Data

Following Gayle and Miller (2009) the price of a bond is computed as the present value (in real terms) of a security (T-bill) which pays \$1 annually. Denote r_{it} the marginal annuitized yield from lengthening the bond one period by extending the maturity date from $t+i$ to $t+i+1$. Data comes from the Federal Reserve's Economic Research Data Base and is based on Treasury bills with maturities 1, 2, 3, 5, 7, 10, 20, and 30. Assume the marginal annuitized yield rate for any bond maturing over 30 years is the same as the 30-year rate.

[65] Relevant data on income and occupation is only collected for household heads.

This yields b_t defined as

$$\begin{aligned} b_t &\equiv \sum_{s=1}^{\infty}\prod_{i=1}^{s}(1+r_{it})^{-1} \\ &= \sum_{s=1}^{30}\prod_{i=1}^{s}(1+r_{it})^{-1}+\prod_{i=1}^{30}(1+r_{it})^{-1}\sum_{s=31}^{\infty}(1+r_{30,t})^{s-30} \\ &= \sum_{s=1}^{30}\prod_{i=1}^{s}(1+r_{it})^{-1}+\frac{1}{r_{30,t}}\prod_{i=1}^{30}(1+r_{it})^{-1} \end{aligned} \tag{20}$$

Then, for each date t, impute a yield curve using the data on newly issued bonds for various maturities. Then use a cubic spline for each date-maturity combination in the data to obtain imputations $\hat{r}_{it}$ for each date t and for all $i \in \{1,\ldots,30\}$.

Step 1:

Use the annual compounding interest rate $\widetilde{r}_{st}$ (from the interpolated yield curve) to obtain b_t as

$$b_t = \sum_{s=1}^{30}\left(\frac{1}{1+\widetilde{r}_{st}}\right)^{s}+\frac{1}{r_{30,t}}\left(\frac{1}{1+\widetilde{r}_{30,t}}\right)^{30} \tag{21}$$

Step 2:

Given that r_{it} and $\widetilde{r}_{it}$ are nominal interest rates, b_t is adjusted by the deflator based on base year 2000. To reflect inflation let

$$\widetilde{b}_t = \frac{b_t}{deflator2000} \tag{22}$$

The series of $\widetilde{b}_t$ is the one used in estimation. Given the sample, the earliest bond price needed is for year 1968 and the last bond price needed is for year 2033. The last yield curve available is for year 2015. Hence in-sample bond prices can be obtained up to 2015. Given the bond prices in sample $\widetilde{b}_t$ for $t = 1954,\ldots,2015$, a regression is run for $\widetilde{b}_{t+1}$ on $\widetilde{b}_t$ and in order to obtain out-of-sample prices $\hat{b}_t$ for $t = 2016,\ldots,2033$.

B Model Appendix

B.1 Claim about belief variance.

Claim: For any $t > t_{i0}$ the prior variance $\mathbb{V}_{it}$ is a deterministic function of the parameters of the population distribution variance matrix $\mathbf{\Delta}$ and the experience vector x_{it}.

Proof: wlog let $t_{i0} = 1$ and drop the i index. First, let $\tilde{\mathbf{\Delta}} = \mathbf{\Delta}^{-1}$ with characteristic component $\tilde{\delta}_{k,k'}$. For $t = 2$, after choosing occupation $j \in \{1, \ldots, 4\}$, from the updating rule in equation (7) we know that

$$\mathbb{V}_2^{-1} = \mathbf{\Delta}^{-1} + \Sigma_1 \tag{23}$$

where Σ_t is defined in (5). The off-diagonal components of $\mathbb{V}_2^{-1}$ are simply $\tilde{\delta}_{k,k'}$. The diagonal components can be written as $\tilde{\delta}_{k,k} + d_{k1}/\sigma^2_{\eta_k} = \tilde{\delta}_{k,k} + x_{k2}/\sigma^2_{\eta_k}$ characteristic components. Now suppose it is also true that the off-diagonal components of $\mathbb{V}_t^{-1}$ are $\tilde{\delta}_{k,k'}$ and the diagonal components are $\tilde{\delta}_{k,k} + x_{kt}/\sigma^2_{\eta_k}$. Then, using the updating rule in equation (7) again we obtain that

$$\mathbb{V}_{t+1}^{-1} = \mathbb{V}_t^{-1} + \Sigma_t \tag{24}$$

From the previous equation it is clear that the diagonal and off-diagonal components of $\mathbb{V}_{t+1}^{-1}$ can also be written as deterministic functions of the components of $\mathbf{\Delta}^{-1}$ and the vector of experience, x_{it+1}. An induction argument finishes the proof. *Q.E.D.*

B.2 Proof of proposition 1

Proof: The proof works by backwards induction. Consider the set up of his problem in the last period of his labor market career, T, in present value terms. Suppose that he has chosen alternative k at period T. His consumption and savings choice maximizes

$$\begin{aligned} - \quad & \alpha_{Tk}(h_T)\beta^T \exp\{-\rho c_T - \varepsilon_{Tk}\} - E_T\left[\lambda_{\tau(T+1)} b_{\tau(T+1)} v_{kT+1} \exp\left(\frac{-(\rho\xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right)\middle| \mathbb{E}_T, h_T\right] \\ & s.t.\ \ E_T[\lambda_{\tau(T+1)}\xi_{T+1} | d_{Tk}, h_T, \mathbb{E}_T] + \lambda_{\tau(T)} c_T = \lambda_{\tau(T)}\xi_T \end{aligned} \tag{25}$$

His budget constraints shows the relation between the value of his wealth today, his consumption choice, and the expected value of his wealth tomorrow. If he works in occupation k he obtains income $\bar{L}_k y_{kt+1}$ at the beginning of his retirement age which is simply added to his wealth in equation (9). Following a similar procedure as in Margiotta and Miller (2000,

p. 680) the conditional value function of choosing alternative k is obtained as

$$V_{kT}(h_T, \mathbb{E}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}, \varepsilon_{kT}) =$$
$$-\lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} e^{-\varepsilon_{kT}/b_{\tau(T)}} E_T[v_{kT+1}|\mathbb{E}_T]^{1-1/b_{\tau(T)}} \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) \tag{26}$$

Integrating over ε_T and averaging over the 5 choices using the conditional choice probabilities yields

$$\begin{aligned} & V_T(h_T, \mathbb{E}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}) = \\ & \quad -\sum_{k=0}^{4} p_{kT}(h_T, \mathbb{E}_T) \lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} E_\varepsilon[e^{-\varepsilon^*_{kT}/b_{\tau(T)}}] E_T[v_{kT+1}|\mathbb{E}_T]^{1-1/b_{\tau(T)}} \\ & \quad \times \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) \\ = & \quad -\lambda_{\tau(T)} b_{\tau(T)} \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) A_T(h_T, \mathbb{E}_T) \end{aligned} \tag{27}$$

where

$$E_\varepsilon[e^{-\varepsilon^*_{kT}/b_{\tau(T)}}] \equiv E_\varepsilon[e^{-\varepsilon_{kT}/b_{\tau(T)}}|d_{kT} = 1]$$

and $A_T(h_T, \mathbb{E}_T)$ is defined as in equation (11) with $A_{T+1}(h_{T+1}, \mathbb{E}_{T+1}) \equiv 1$.

To finish the proof suppose that equations (10) and (11) hold for $t+1$. Then, at age t an individual who has chosen alternative k selects consumption and savings to maximize

$$\begin{aligned} - & \quad \alpha_{kt}(h_t)\beta^t \exp\{-\rho c_t - \varepsilon_{kt}\} \\ - & \quad E_t\left[\lambda_{\tau(t+1)} b_{\tau(t+1)} A_{t+1}(h_{t+1}, \mathbb{E}_{t+1}) v_{kt+1} \exp\left(\frac{-(\rho\xi_{t+1} + a_{\tau(t+1)})}{b_{\tau(t+1)}}\right) \middle| \mathbb{E}_t, h_t, d_{kt} = 1\right] \\ & \quad s.t. \;\; E_t[\lambda_{\tau(t+1)}\xi_{t+1}|d_{kt}, h_t, \mathbb{E}_t] + \lambda_{\tau(t)} c_t = \lambda_{\tau(t)}\xi_t \end{aligned}$$

Which yields an equation similar to equation (26):

$$
\begin{aligned}
&V_{kt}(h_t, \mathbb{E}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}, \varepsilon_{kt}) = \\
&-\lambda_{\tau(t)} b_{\tau(t)} \alpha_{kt}(h_t)^{1/b_{\tau(t)}} e^{-\varepsilon_{kt}/b_{\tau(t)}} E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1} | \mathbb{E}_t, h_t]^{1-1/b_{\tau(t)}} \\
&\times \exp\left(\frac{-(\rho\xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right) \qquad (28)
\end{aligned}
$$

The proof is finished by integrating over ε_t and averaging over the 5 choices using the conditional choices probabilities. *Q.E.D.*

B.3 Proof of Proposition 2

Proof: Assuming that the taste shocks are distributed Extreme Value Type-I renders the expression in equation (12) as a standard logit. Hence, the odds ratio can be written as

$$
\frac{p_{0t}(h_t, \mathbb{E}_t)}{p_{kt}(h_t, \mathbb{E}_t)} = \alpha_{kt}(h_t) E_t \left[\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})}{A_{t+1}(h_t, \mathbb{E}_t)} v_{kt+1} \middle| \mathbb{E}_t, h_t \right]^{b_{\tau(t)}-1} \qquad (29)
$$

Equation (29) describes the likelihood ratio of any choice relative to the choice of not working. The reason why the arguments of the index in the denominator are subscripted with t is that neither the individual's human capital vector nor his beliefs change if he decides not to work. Use equation (29) to write

$$
E_t \left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1} \middle| \mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}} = \alpha_{kt}(h_t)^{-1/b_{\tau(t)}} A_{t+1}(h_t, \mathbb{E}_t)^{1-1/b_{\tau(t)}} \left(\frac{p_{kt}(h_t, \mathbb{E}_t)}{p_{0t}(h_t, \mathbb{E}_t)}\right)^{-1/b_{\tau(t)}} \qquad (30)
$$

From page 3 in the online appendix of Gayle et al. (2015):

$$
E_\varepsilon[e^{-\varepsilon^*_{kt}/b_{\tau(t)}}] = p_{kt}(h_t, \mathbb{E}_t)^{1/b_{\tau(t)}} \Gamma\left(\frac{b_{\tau(t)}+1}{b_{\tau(t)}}\right) \qquad (31)
$$

where $\Gamma(\cdot)$ denotes the complete gamma function. Substitute equations (30) and (31) in equation (11) to obtain

$$
A_t(h_t, \mathbb{E}_t) = p_{0t}(h_t, \mathbb{E}_t)^{1/b_{\tau(t)}} \Gamma\left(\frac{b_{\tau(t)}+1}{b_{\tau(t)}}\right) A_{t+1}(h_t, \mathbb{E}_t)^{1-1/b_{\tau(t)}} \qquad (32)
$$

Using equation (32) we can write the ratio of human capital and beliefs indexes as

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})}{A_{t+1}(h_t, \mathbb{E}_t)} = \frac{p_{0t+1}(h_{kt}^{(1)}, \mathbb{E}_{kt}^{(1)})^{1/b_{\tau(t)+1}} A_{t+2}(h_{kt}^{(1)}, \mathbb{E}_{kt}^{(1)})^{1-1/b_{\tau(t)+1}}}{p_{0t+1}(h_{0t}^{(1)}, \mathbb{E}_{0t}^{(1)})^{1/b_{\tau(t)+1}} A_{t+2}(h_{0t}^{(1)}, \mathbb{E}_{0t}^{(1)})^{1-1/b_{\tau(t)+1}}} \tag{33}$$

where $h_{kt}^{(1)}$ and $\mathbb{E}_{kt}^{(1)}$ indicate the value of the state variables at future age $t+1$, conditional on the decision path described by making $d_{kt} = 1$. In general, define $h_{kt}^{(s)}$ and $\mathbb{E}_{kt}^{(s)}$ as the value of the state variables at future age $t+s$, conditional on the decision path described by making $d = 1$ for all $d \in \{d_{kt}, d_{0t+1}, \ldots, d_{0T}\}$ and define

$$\phi_t(s) = \frac{1}{b_{\tau(t)+s}} \prod_{r=1}^{s-1} (1 - 1/b_{\tau(t)+r}) \tag{34}$$

Iterative substitution of equation (32) in (33) up to retirement age yields

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})}{A_{t+1}(h_t, \mathbb{E}_t)} = \prod_{s=1}^{T-t} \left(\frac{p_{0t+s}(h_{kt}^{(s)}, \mathbb{E}_{kt}^{(s)})}{p_{0t+s}(h_{0t}^{(s)}, \mathbb{E}_{0t}^{(s)})} \right)^{\phi_t(s)} \tag{35}$$

Plugging equation (35) into equation (29) and applying logarithms finishes the proof. *Q.E.D.*

C Estimation Appendix

C.1 First Stage Detailed

The first stage uses an Expectation-Maximization algorithm. The EM algorithm is an iterative method that yields maximum likelihood estimates when a portion of the relevant data is unobserved. In the model, the unobserved part of the data is $\mathcal{M}_i$. In order to implement the EM algorithm assume $\mathcal{M}_i$ is observed for all i. Hence, the log likelihood of the data for individual i is

$$\begin{aligned}\ln \mathcal{L}_i &= \sum_{t=t_{i0}}^{T_i}\sum_{k=0}^{4} d_{kit} \ln \Pr\left[d_{kit}=1|h_{it},\mathbb{E}_{it};\Lambda,\Theta\right] \\ &\quad + \sum_{t=t_{i0}}^{T_i}\sum_{j=0}^{4} d_{jit} \ln \Pr\left[y_{jit+1}|h_{it},\mu_j;\Theta\right] \\ &\equiv \ln \mathcal{L}_i^P + \ln \mathcal{L}_i^y \qquad (36)\end{aligned}$$

Since $\ln \mathcal{L}_i$ is additively separable, $\ln \mathcal{L}_i^y$ is used to consistently estimate Θ and $\mathbf{\Delta}_s$ using the EM algorithm. Implementation of the EM algorithm iterates over two steps to obtain maximum likelihood estimates. The expectation step at the mth iteration requires computation of the expectation of $\ln \mathcal{L}_i^y$ conditional on the observed data and the parameters at the mth iteration. The maximization step finds the new iterated value of the vector of parameters by maximizing the expression obtained in the expectation step.

Expectation Step. Using Bayes' rule (see Ch. 9 in DeGroot (1970) and James (2011)), the conditional distribution of $\mathcal{M}_i$ for an individual with education level s at the m^{th} iteration, based on the observed data, is $N\left(\mathbb{E}_i^m,\mathbb{V}_i^m\right)$ where

$$\mathbb{E}_i^m = \left(\left(\mathbf{\Delta}_s^m\right)^{-1}+\Psi_i\right)^{-1}\mathbf{W}_i \qquad (37)$$

$$\mathbb{V}_i^m = \left(\left(\mathbf{\Delta}_s^m\right)^{-1}+\Psi_i\right)^{-1} \qquad (38)$$

where the kth component of the $\mathbf{W}_i$ vector is

$$\mathbf{W}_{i\{k\}} = \frac{\sum_{t=1}^{T} d_{kit}\left(y_{kit}-h_{it}'\theta_k\right)}{\sigma_k^{2,m}}$$

and the diagonal components of the square matrix Ψ_i are

$$\Psi_{i\{k,k\}} = \frac{\sum_{t=1}^{T} d_{kit}}{\sigma_k^{2,m}}$$

The off-diagonal terms of Ψ_i are all zeros. Given μ_{ki} and the distribution of η_{kit}

$$\begin{aligned}
\log \Pr\left[y_{it}|h_{it},\mu_k;\Theta\right] &= \log\left(\frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left\{\frac{-\left(y_{kit}-h'_{it}\theta_k-\mu_{ki}\right)^2}{2\sigma_k^2}\right\}\right) \\
&= -\frac{1}{2}\log\left(2\pi\sigma_k^2\right)-\frac{1}{2\sigma_k^2}\left(y_{kit}-h'_{it}\theta_k-\mu_{ki}\right)^2
\end{aligned}$$

Therefore, the expectation step of the EM algorithm yields

$$\begin{aligned}
E_m\left[\log\mathcal{L}_i^y\right] &= -\sum_{t=1}^{T}\sum_{k=1}^{4} d_{kit}\cdot E_m\left[\frac{1}{2}\log\left(2\pi\sigma_k^2\right)+\frac{1}{2\sigma_k^2}\left(y_{kit}-h'_{it}\theta_k-\mu_{ki}\right)^2\right] \\
&= -\sum_{t=1}^{T}\sum_{k=1}^{4} d_{kit}\left[\frac{1}{2}\log\left(2\pi\sigma_k^2\right)+\frac{1}{2\sigma_k^2}\left(\mathbb{V}_{i\{k,k\}}^m+\left(y_{kit}-h'_{it}\theta_k-\mathbb{E}_{i\{k\}}^m\right)^2\right)\right]
\end{aligned} \tag{39}$$

where $E_m[\cdot]$ stands for the expectation over $\mathcal{M}_i$ using the distribution characterized by the parameters of the mth iteration conditional on the observed data.

Maximization Step. Following the expectation step, the maximization step entails maximizing (39) in order to obtain Θ^{m+1}. In fact, each θ_k^{m+1} is given by

$$\theta_k^{m+1} = \arg\min_{\theta_k}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{kit}\left(y_{it}-h'_{it}\theta_k-\mathbb{E}_{i\{k\}}^m\right)^2 \tag{40}$$

which yields

$$\theta_k^{m+1} = (H'W_kH)^{-1}H'W_kY_k$$

where H is the $[NT\times\#(\theta_k)]$ matrix that stacks together all values of h'_{it}, Y_j is the $[NT\times 1]$ matrix that stacks together all values of $y_{it}-\mathbb{E}_{i\{k\}}^m$, and W_k is the $[NT\times NT]$ diagonal matrix with d_{kit} in its diagonal. Using the FOCs from (39) and the estimated values of θ_j^{m+1}, $\sigma_k^{2,m+1}$

has the closed form solution

$$\sigma_k^{2,m+1} = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} d_{kit}\left(\mathbb{V}_{i\{k,k\}}^m + \left(y_{it} - h_{it}'\theta_k^{m+1} - \mathbb{E}_{i\{k\}}^m\right)^2\right)}{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} d_{kit}} \tag{41}$$

A summary of the EM algorithm is

↪ Step 1: Given mth iteration values $\left\{\theta_k^m, \sigma_k^{2,m}\right\}_{k\in\{1,\dots,4\}}$ and $\{\mathbf{\Delta}_s^m\}_{s\in\{1,\dots,4\}}$, solve for $\mathbb{E}_i^m$ and $\mathbb{V}_i^m$ using (37) and (38).

↪ Step 2: Update population parameter $\mathbf{\Delta}_s^{m+1}$ for education level s as

$$\mathbf{\Delta}_s^{m+1} = \frac{1}{N_s}\sum_{i=1}^{N}\sum_{s=1}^{4}\delta_{is}\left(\mathbb{V}_i^m + \mathbb{E}_i^m\mathbb{E}_i^{m\prime}\right) \tag{42}$$

where δ_{is} is an indicator of individual i having education level s and $N_s = \sum_i \delta_{is}$. Equation (42) follows from maximization of the expected value of the log likelihood of $\mathcal{M}_i$, $E_m\left[\log f(\mathcal{M}_i)\right]$.[66]

↪ Step 3: For each occupation $k > 0$, new iteration values θ_k^{m+1} are obtained using equation (40) and new iteration values $\sigma_k^{2,m+1}$ are obtained using equation (41).

The algorithm is initialized with arbitrary values and the steps are repeated until convergence under the criterion

$$\left\|\sum_{i=1}^{N}\log\tilde{\mathcal{L}}_i^{y,m+1} - \sum_{i=1}^{N}\log\tilde{\mathcal{L}}_i^{y,m}\right\| < \epsilon$$

where

$$\tilde{\mathcal{L}}_i^{y,m} = \int\limits_{\tilde{\mathcal{M}}}\left\{\prod_{t=t_{i0}}^{T_i}\prod_{j=1}^{4}\Pr\left[y_{jit+1}|h_{it},\tilde{\mu}_j;\Theta^m\right]^{d_{jit}}\right\} dF\left(\tilde{\mathcal{M}};\mathbf{\Delta}_s^m\right) \tag{43}$$

is computed using Monte Carlo integration. ϵ is set to be 1×10^{-4}

C.2 Second Stage Detailed

The second stage of the estimation procedure is initialized with flexible parametric versions of the future conditional choice probabilities estimated from the data, where the beliefs,

[66]See Anderson and Olkin (1985).

estimated in the first stage, are also treated as data. In the model, individuals have perfect foresight about their marital status. However, their entire marital status vector up to period T is not always observed. Hence, their marital status histories are completed using a single marital status path constructed using the median age of first marriage at 1970 from the U.S. Census Bureau, Current Population Survey and the median marriage duration presented in Kreider and Ellis (2011).[67] Effectively it amounts to individuals getting married at age 23 and remaining married until age 50.

C.2.1 Maximization Step

At any iteration of the second stage, for a given set of estimated ccps, utility parameters are obtained from maximization of the log likelihood

$$\frac{1}{NT}\sum_{i}\sum_{t}\sum_{k=0}^{4} d_{kit}\ln p_{kit}(h_t,\mathbb{E}_t) \tag{44}$$

The expectation in the expression for $V_k(h_t,\mathbb{E}_t)$ in equation (17) can be written as

$$\begin{aligned}
&E_t\left[v_{kit+1}\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)},\mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)},\mathbb{E}_{0it}^{(s)})}\right)^{\phi(s)}\middle| \mathbb{E}_{it},h_{it}\right]\\
&=\int_{\zeta_k}\left\{\exp\left(\frac{-\rho\bar{L}_k y_{kit+1}(h_{it})}{b_{\tau(t+1)}}\right)\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)},\mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)},\mathbb{E}_{it})}\right)^{\phi(s)}\right\}dF(\zeta_k|\mathbb{E}_{it},h_{it})\\
&=\int_{\zeta_k}\left\{\exp\left(\frac{-\rho\bar{L}_k(f_k(h_{it},\omega_i;\theta_k)+\zeta_k)}{b_{\tau(t+1)}}\right)\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)},\mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)},\mathbb{E}_{it})}\right)^{\phi(s)}\right\}dF(\zeta_k|\mathbb{E}_{it},h_{it})
\end{aligned} \tag{45}$$

I compute the value of (45) using Monte Carlo integration.

Given a value for ρ the model becomes a simple logit in the α parameters and the Monte Carlo integral is[68]

$$B_{kit}(\rho)=\frac{1}{S}\sum_{s}\left\{\exp\left(\frac{-\rho\bar{L}_k(f_k(h_{it};\theta_k)+\zeta_k^s)}{b_{\tau(t+1)}}\right)A_{kit}(\zeta_k^s)\right\} \tag{46}$$

[67] For the median age at first marriage visit: `http://www.census.gov/hhes/families/data/marital.html`.

[68] Recall that the scale of parameters Θ, $\boldsymbol{\Delta}_s$, and ρ depends on the units in which income and consumption are measured. I express hourly income in \$10 units and consumption in \$1,000 units. Therefore, in estimation, instead of $\bar{L}_k$ I write $\bar{L}_k/100$.

where

$$A_{kit}(\zeta_k^s) = \prod_{s=1}^{T-t} \left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k^s))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi(s)}$$

is a value which varies across signals, ζ_k^s, drawn for integration. The draws come from the distribution of the signal conditional on beliefs $\zeta_k^s = \mu_{ki} + \eta_{kit} \sim N\left(\mathbb{E}_{\{k\}it}, \mathbb{V}_{\{k,k\}it} + \sigma_{\eta_k}^2\right)$. Using the definition of $\alpha_{kt}(h_t)$ in equation (3), equation (17) can be rewritten as

$$V_k(h_t, \mathbb{E}_t) = -h_{it}'\alpha_k - C_{kit}(\rho) \tag{47}$$

where

$$C_{kit}(\rho) = (b_{\tau(t)} - 1) \ln B_{kit}(\rho)$$

Equation (47) is then substituted into (16). In estimation, the log likelihood is maximized conditional on a value of ρ. Search over values for ρ is then undertaken an the value that maximizes the log likelihood is selected. This procedure is faster that searching over all the parameter space at once because it avoids computing the Monte Carlo integral in (45) more than once for each value of ρ.

C.2.2 CCP Step

For a given value of utility parameters the model is solved backwards and new model-generated ccps are obtained. This new ccps are fed into the maximization step and new utility parameters are obtained. Given that each iteration is computationally intensive, the iterative process is stopped after 5 iterations because the minimum log likelihood is achieved in iteration 4. The Euclidean distance between the parameter vectors in iteration 4 and 5 is 9.4. Future version of the model may entail more iterations to ensure that the solution is not a local minimum. The relatively small distance between the parameter vectors and the fact that the first search was initialized from 10 different initial points suggests that the solution may not be local. Notice that the estimated parameters at each iteration are consistent since the ccps that initialize the process are themselves consistent.

D Results Appendix

D.1 Solving the Model

As mentioned in the Estimation section, solution of the model is needed in order to provide new estimates of the conditional choice probabilities. The model is solved using the same representation obtained in Proposition 2 and summarized in equation (17). Notice that this representation is obtained as a function of the probability of not working in the future conditional on specific choice paths. Consistent, with this representation, for a given vector of estimated parameters, the value function is solved with the following recursive algorithm starting at $t = T$:

- ↪ *Step 1.* Obtain the value of the mapping $V_k(h_{it}, \mathbb{E}_{it})$ for a grid spanning the relevant state space using equation (17) and the future choice paths described in Proposition 2.[69]
- ↪ *Step 2.* Obtain relevant ccps for period t using equation (16).
- ↪ *Step 3.* Obtain parametric versions of the ccps for period t, Ω_t. Noting that only the not working ccps are needed, the parametric version is obtained using a non linear regression that minimizes the distance between the model ccps, $p_{0it}(h_{it}, \mathbb{E}_{it})$, from Step 2 and a parameterization given by $\exp(X_{it}'\Omega_t)/(1 + \exp(X_{it}'\Omega_t))$. X_{it} includes multiple interactions of components of the state.
- ↪ *Step 4.* If $t = t_0$, stop. Otherwise, set $t = t - 1$ and go back to Step 1 using the collection of parametric ccps obtained so far for the representation.

As a final product from the previous algorithm, a collection of future ccps, $\{\Omega_t\}_{t=t_0}^{T}$, characterizing the value function at any period t is obtained.

D.2 Model Fit

In order to assess goodness of fit, an initial state is generated and the model is simulated forward using the collection of future ccps implied by the model, $\{\Omega_t\}_{t=t_0}^{T}$, that characterize the value function. For comparison against the data initial states are obtained drawing from the data under certain restrictions. First, a collection of initial states is drawn from the initial states observed in the data (race, education, entry age, year of entry, permanent

[69]Notice that at period T there is no future value of human capital and beliefs. Only the value of income to be received at $T + 1$. Hence, no future ccps are needed as the occupational choice becomes static.

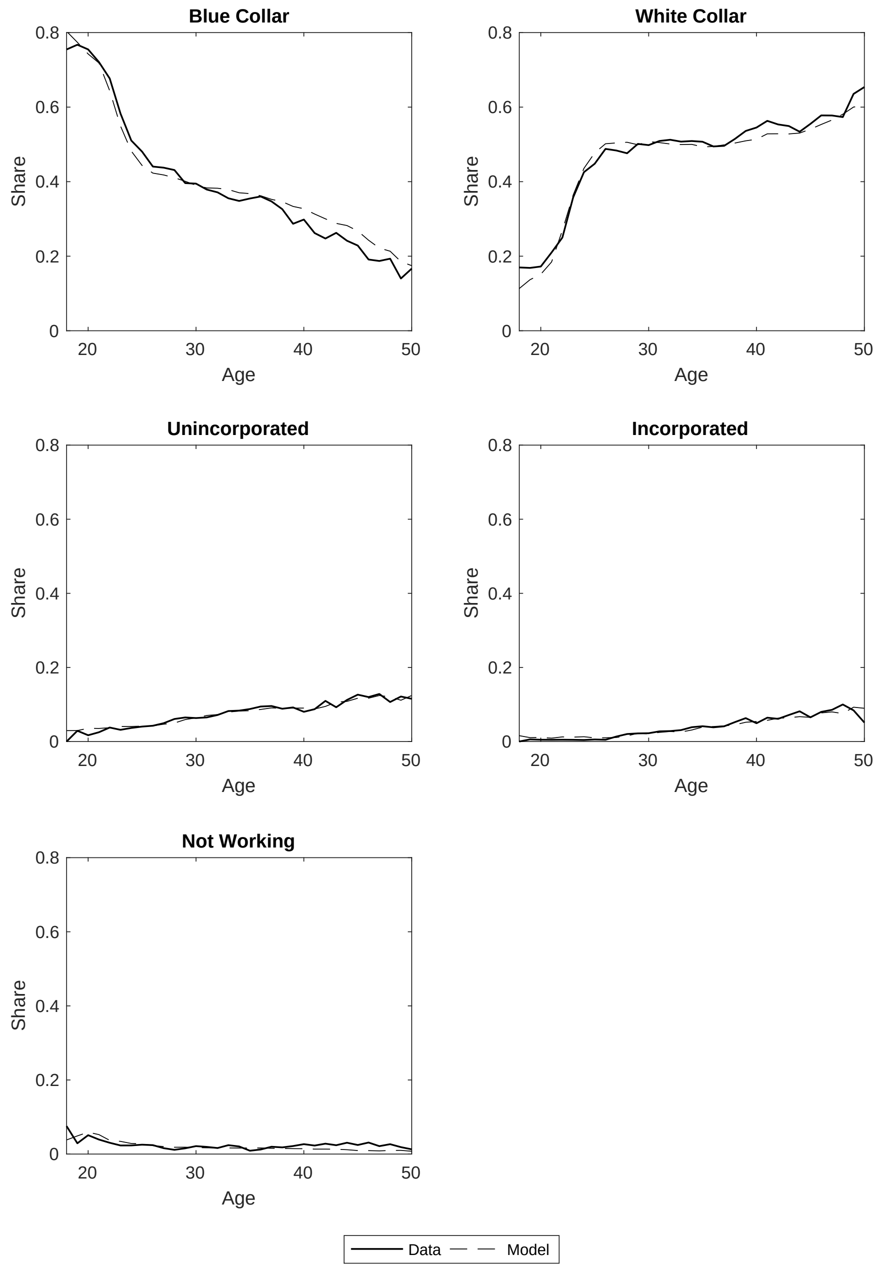
wealth). To avoid the high volatility of bond prices before 1980, only years after 1980 are considered for the comparison. Second, only one marital status path is allowed: the one constructed in the Estimation Appendix. Third, ability for each individual drawn from the data is set to be the mean beliefs conditional on all the information available for him. Using Bayes' rule, this is equal to his beliefs in the last period the individual is observed. In order to increase precision, only individuals that are observed for at least ten years are used in the comparison against the data.

An initial measure of model fit is presented in Figure 18. It shows that the estimated model replicates the choices well given the observed state. Table 15 compares the transition matrices. In the model, occupations are less absorbing than in the data. However, consisntent with the data, entrepreneurial occupations are on average less sticky than salaried occupations. Notably, the not working alternative is much less absorbing in the model, which suggests that there are barriers to exit unemployment that are not captured in the model. In terms of switching behavior, the model successfully captures the fact that most switching from salaried occupations happens within the salaried group. It also captures the fact that, whereas unincorporated individuals tend to switch in similar percentages into either salaried occupation, incorporated entrepreneurs tend to overwhelmingly switch into white collar work. Table 16 compares descriptive statistics of occupational spells. Although, consistent with transition results, the model under-predict spell durations, it performs well in terms of the distribution of occupational spells across-occupations. At the beginning of their careers, the model over predicts the number of individuals starting as unemployed or blue collar workers, and under predicts the number starting as white collar workers.

.

Table 17 shows the mean and variance of hourly income across all individuals who participate in each of the four occupations. With the exception of incorporated entrepreneurship, the model captures well the first two moments of the income distribution. For incorporated entrepreneurs, the model over predicts mean and variance. Notwithstanding this over prediction, the model respects the relative order in terms of which occupations generate higher income variance and which ones offer higher average income. Additionally, the model is able to capture the trend in the relation between the probability of switching into entrepreneurship from salaried occupations and the current income signal (see Figure 2). Figure 19 shows that better signals in salaried occupations are negatively associated with the probability of switching into unincorporated entrepreneurship. The model captures the relative flatness of the relation between the signal and the probability of switching into incorporated entrepreneurship from blue collar work. Moreover, the increase in the probability of switching into incorporated entrepreneurship from white collar, for those receiving the best signals, is

Figure 18: Model Fit by Age



Notes: Actual and simulated choices by age.

also captured by the model. Consistent with the excess switching shown in the simulated transition matrix, the model is unable to capture the level of the relations found in the data. Highlighting the role of correlated learning, Figure 20 shows that a model that does not allow

Table 15: Transition Patterns: Observed and Simulated

Data

	blue collar	white collar	unincorporated	incorporated	not working
blue collar	0.87	0.08	0.02	0.00	0.02
white collar	0.07	0.89	0.02	0.01	0.01
unincorporated	0.11	0.10	0.73	0.04	0.01
incorporated	0.03	0.15	0.06	0.75	0.01
not working	0.37	0.15	0.03	0.00	0.44

Model

	blue collar	white collar	unincorporated	incorporated	not working
blue collar	0.74	0.18	0.04	0.01	0.03
white collar	0.22	0.71	0.05	0.02	0.01
unincorporated	0.23	0.21	0.53	0.02	0.01
incorporated	0.08	0.17	0.03	0.72	0.00
not working	0.62	0.25	0.05	0.01	0.07

Notes: Matrix entry i, j represents the proportion of people in occupation in row i who move into occupation in column j between t and $t+1$.

Table 16: Spells: Observed and Simulated

Data

	all	blue collar	white collar	unincorporated	incorporated	not working
Total	4294	1707	1652	453	194	288
Percent		39.75	38.47	10.55	4.52	6.71
Duration	4.97	5.21	6.03	3.10	3.10	1.63
First		52.06	42.56	2.19	0.27	2.92
Tried		68.73	69.92	20.05	9.03	14.54

Model

	all	blue collar	white collar	unincorporated	incorporated	not working
Total	282999	114681	104919	34842	11183	17374
Percent		40.52	37.07	12.31	3.95	6.14
Duration	3.03	3.54	3.13	2.02	3.02	1.07
First		56.84	29.02	3.46	0.74	9.94
Tried		77.06	76.74	32.04	16.44	39.99

Notes: **Duration** is the average duration of spells in years. **First** is the percentage of first spells that belong to a particular occupation. **Tried** is the percentage of individuals who tried the occupation during their observed careers.

for correlated learning is unable to capture neither the level not the trend of this relation.

D.3 Certainty Equivalent

D.3.1 Static

In order to get a sense of the magnitude of the estimated risk aversion parameter consider a static individual with beliefs $\mathbb{B}_t = \{\mathbb{E}_t, \mathbb{V}_t\}$. His expected annual income from working in

Table 17: Income: Observed and Simulated

Data

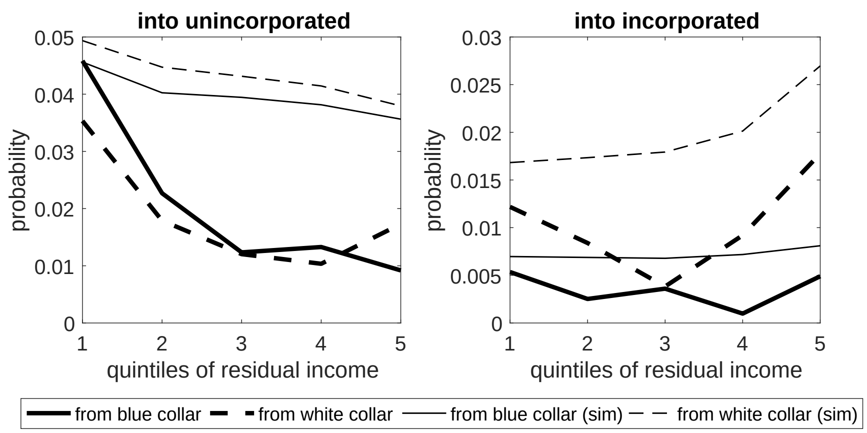
	blue collar	white collar	unincorporated	incorporated
mean income	14.14	21.17	21.00	37.48
variance income	7.94	14.25	22.77	51.17

Model

	blue collar	white collar	unincorporated	incorporated
mean income	15.21	22.35	23.88	51.04
variance income	5.83	13.04	27.47	89.57

Notes: Quantities in dollars of 2000.

Figure 19: Probability of Switching into Entrepreneurial Occupations



Notes: Probability of switching into entrepreneurial occupations in $t+1$ by decile of residual income in t. Residual income is computed from occupation-specific regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status.

occupation k at age t is

$$\bar{y}_{kt+1} = f_k(h_{it}; \theta_k) + \mathbb{E}_{t\{k\}}$$

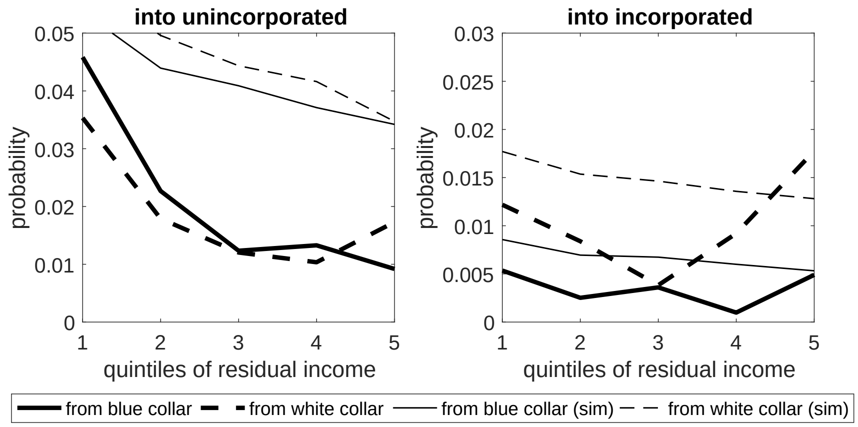
and he considers the variance of his hourly income to be

$$\sigma_{kt}^2 = \mathbb{V}_{t\{k,k\}} + \sigma_{\eta_k}^2$$

Therefore, his certainty equivalent at occupation k, y_k^c, solves

$$-\exp\left\{-\rho \bar{L}_k y_k^c\right\} = -\exp\left\{-\rho \bar{L}_k \bar{y}_{kt+1} + \frac{\rho^2 \bar{L}_k^2}{2}\sigma_{kt}^2\right\}$$

FIGURE 20: Probability of Switching into Entrepreneurial Occupations (Uncorrelated Learning)



Notes: Probability of switching into entrepreneurial occupations in $t+1$ by decile of residual income in t under the counterfactual that learning about ability is uncorrelated. Residual income is computed from occupation-specific regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status.

which yields

$$y_k^c = \bar{y}_{kt+1} - \frac{\rho \bar{L}_k}{2} \sigma_{kt}^2 \tag{48}$$

In estimation $\bar{L}_k$ is substituted with $\bar{L}_k/100$.

D.3.2 Dynamic

In order to obtain the dynamic version of the certainty equivalent use equation (28) to find the quantity y_k^c such that

$$E_t \left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}} \right) |\mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}} = E_t \left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t) \exp\left(\frac{-\rho \bar{L}_k y_k^c}{b_{\tau(t+1)}} \right) |\mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}}$$

As opposed to the static case, the future value of human capital and beliefs also determine the certainty equivalent:

$$y_k^c = -\left(\frac{b_{\tau(t+1)}}{\rho \bar{L}_k} \right) \ln \left(\frac{E_t \left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1} |\mathbb{E}_t, h_t \right]}{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t)} \right) \tag{49}$$

D.4 Monetary Value of Entry Costs

The monetary value of entry costs is obtained using equation (28). From equation (3) one can separate the non-pecuniary costs in two factors, one corresponding to the entry costs, $\alpha^e_{kt}(h_t)$ and the other corresponding to all other non-pecuniary costs. Hence, let $\alpha_{kt}(h_t) = \alpha^o_{kt}(h_t) \times \alpha^e_{kt}(h_t)$. Next, use equation (3) to figure out the quantity that should be taken out of annual income in the budget constraint in order to equalize the conditional value functions. In other words, find the quantity ψ that must be given to the individual to leave him indifferent between (a) receiving ψ and facing entry costs and (b) not receiving ψ but facing no entry costs. It solves:

$$\begin{aligned}
&\alpha^e_{kt}(h_t)^{1/b_{\tau(t)}} E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right) |\mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}} \\
&= E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho(\bar{L}_k y_{kt+1}(h_t) - \psi)}{b_{\tau(t+1)}}\right) |\mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}}
\end{aligned}$$

which yields

$$\psi = \frac{\ln \alpha^e_{kt}(h_t)}{\rho} \frac{b_{\tau(t+1)}}{b_{\tau(t)} - 1} \tag{50}$$

Since the quantity $\bar{L}_k y_{kt+1}$ was written in thousands of dollars in estimation, the value of ψ is in thousands of dollars.

D.5 Alternative Regimes

In order to increase precision and facilitate comparison across alternatives, in this section ability is not approximated from the data. Instead, individuals' ability vectors are drawn from the estimated distributions in Table 7. Rather than being replicated from the data, individuals are simulated using the empirical joint distribution of initial states. Simulations are undertaken using a fictional economy in which there is no aggregate variation in bond prices. In this stationary environment the bond price is set to remain constant at the 1990 level.[70] Marital status paths follow the same restriction specified in the Estimation Appendix. The initial state and bond price sequence used in the decomposition is also used for the policy counterfactuals.

These counterfactual regimes are described below:

[70] An alternative way of dealing with the aggregate variation is to undertake a partial equilibrium analysis that fixes the sequence of bond prices observed in the data across counterfactual regimes.

↪ C1: *No Learning-by-doing.* In this counterfactual, individuals receive a fix hourly return regardless of how much experience in the occupation they have accumulated. The fixed hourly return provided to individuals is constructed as an approximation of the average returns from experience in the occupation. It is computed as the average of the returns to experience, computed during the first 20 years in the labor market, of an individual that works exclusively in the occupation. Let $R_k(x)$ be the returns to experience in occupation k for somebody that has worked x years in occupation k and zero years in any other occupation (see Figure 4). Then, the fixed hourly return to individuals in occupation k under this counterfactual is

$$\bar{y}_k = \sum_{x=0}^{20} R_k(x)$$

This exercise yields the following values:

TABLE 18: Average Income for Counterfactual: No Learning-by-doing

	blue collar	white collar	unincorporated	incorporated
$\bar{y}_k$	0.378	0.879	0.725	1.234

Notes: Computed using the profiles in Table 6.

Individuals under this counterfactual still have differential returns based on their education, race, marital status, and ability.

↪ C2: *Isolated full information about ability.* In this counterfactual, individuals have full information about their ability. In order to isolate the effect of sorting on ability, the value of the idiosyncratic income variance of their income is set to equal its original value (see Table 8) plus the value of the ability variance (see Table 7). In terms of equation (4), this amounts to changing the value of the idiosyncratic income variance in occupation k from just σ_{η_k} to $\sigma_{\eta_k} + V_{\{k,k\}}$.

↪ C3: *No Cross-occupation Returns.* In this counterfactual, the returns in occupation k from experience accumulated in occupation $k' \neq k$ (see Figure 5) are set to be zero.

↪ C4: *Uncorrelated Learning.* In this counterfactual, individuals use an alternative variance-covariance matrix in order to update their beliefs. This variance-covariance is formed setting at zero the off-diagonal terms of the variance-covariance matrix of the distribution of ability.

- ↪ C5: *No Uncertainty.* In this counterfactual, individuals have full information about their ability and they face no extra uncertainty coming from the idiosyncratic variance. In other words, this is the same as counterfactual C2 plus setting the idiosyncratic variance to zero.

- ↪ C6: *Uniform Entry Costs.* In this counterfactual, individuals of all ages pay the same entry cost, provided they have the same permanent wealth. This cost equals the one faced by a 35 year old individual with their education level (see Table 9).

Table 19: Comparison of Counterfactual Regimes

Unincorporated

	Baseline	C1	C2	C3	C4	C5	C6
Ever tried	0.31	0.38	0.34	0.28	0.32	0.35	0.54
Ever tried in first 5 years	0.08	0.08	0.11	0.06	0.08	0.12	0.31
PVI if ever tried	518	510	663	493	481	666	618
Spell duration	2.17	2.16	2.81	2.37	2.07	2.89	3.10
Participation rate at age 40	0.10	0.13	0.14	0.10	0.10	0.15	0.26
At first entry							
Ability (10$ per hour)	0.05	0.05	0.59	0.09	-0.09	0.57	0.03
Belief (10$ per hour)	0.04	0.04	-	0.07	0.00	-	0.01
Age	34.07	34.88	32.80	34.91	33.99	32.53	28.04
exp_{bc}	6.76	8.37	6.01	9.76	6.47	5.90	0.85
exp_{wc}	4.77	3.86	4.25	2.55	5.09	4.04	3.88
Overall							
Ability (10$ per hour)	0.37	0.39	1.16	0.45	0.21	1.12	0.23
College or more	0.55	0.58	0.65	0.63	0.49	0.65	0.44

Incorporated

	Baseline	C1	C2	C3	C4	C5	C6
Ever tried	0.15	0.33	0.20	0.14	0.13	0.26	0.26
Ever tried in first 5 years	0.02	0.03	0.05	0.01	0.02	0.07	0.12
PVI if ever tried	757	624	1115	622	576	1115	840
Spell duration	2.88	2.98	5.11	2.91	2.62	5.57	3.39
Participation rate at age 40	0.04	0.09	0.10	0.04	0.03	0.14	0.13
At first entry							
Ability (10$ per hour)	0.55	0.43	1.73	0.50	-0.25	1.52	0.20
Belief (10$ per hour)	0.64	0.49	-	0.54	0.00	-	0.26
Age	38.62	40.52	36.06	40.59	38.74	35.43	29.19
exp_{bc}	6.95	11.43	5.98	13.50	7.03	6.07	0.45
exp_{wc}	8.14	5.68	6.52	3.97	8.09	5.85	5.18
Overall							
Ability (10$ per hour)	1.18	0.92	2.70	1.12	0.24	2.37	0.42
College or more	0.70	0.62	0.66	0.64	0.61	0.63	0.37

Notes: Average of several summary statistics across alternative regimes. *Rows*: **PVI** stands for the present value of income in thousands of dollars. This average is computed only over those who tried the occupation. **At first entry** indicates that quantities are computed at first entry. **Ability** contains the ability of those entering the occupation. **Belief** contains the mean of the belief regarding ability. Both ability and the mean of the belief are in 10$ per hour. $\mathbf{exp_{bc}}$ and $\mathbf{exp_{wc}}$ stand for blue and white collar experience. **Overall** indicates that quantities are computed across all observations of individuals participating in the occupation. *Columns*: **Baseline** is the model specification used in the paper. Columns C4 to C7 correspond to the solution and simulation of the model under alternative regimes. **C1** shuts down accumulation of human capital through experience. All individuals going into occupation k receive the equivalent of the average return from experience of somebody who always works in occupation k. The average is computed over the first 20 years of his labor market career. **C2** is a full information model where the overall level of initial uncertainty is maintained in order to isolate the effect of sorting on ability from risk aversion. In this counterfactual, the idiosyncratic variance is set to be $\sigma_{\eta_k} + V_{\{k,k\}}$. **C3** sets the cross-occupation returns to experience to be zero. **C4** shuts down correlated learning. **C5** is the full information model without uncertainty. In this counterfactual, the idiosyncratic variance σ_{η_k} is set to be zero. **C6** keeps the entry costs constant relative to age. Entry costs are always those of a 35 years old person. However, entry costs still vary with permanent wealth. The same simulated individuals, including their ability vector, is keep constant across counterfactual regimes.