

# Entrepreneurship over the Life Cycle: Where are the Young Entrepreneurs?

Andrés Hincapié<sup>1</sup>

October 12, 2018

## Abstract

Most individuals do not start a business and, if they do, they start well into their thirties. I study multiple mechanisms explaining these stylized facts. Using the Panel Study of Income Dynamics, I estimate a dynamic Roy model with accumulation of experience, risk aversion, and imperfect information about ability. Risk aversion reduces entrepreneurship by up to 40% and providing full information about ability increases it by 35%. The gap in first entry ages between paid employment and entrepreneurship results mainly from entry costs and information frictions. I study counterfactual policies (subsidies and education) that target these barriers to young entrepreneurship, thereby closing the gap, and show that fostering young entrepreneurship can yield higher returns than fostering entrepreneurship later in individuals' careers because the gains from early information are larger.

**KEYWORDS:** Entrepreneurship, young entrepreneurship, occupational choice, experimentation, correlated learning, risk aversion, information value, entrepreneurship education, structural estimation.

---

<sup>1</sup> Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 CB 3305. E-mail: andres.hincapie@unc.edu. I thank my dissertation advisors Barton Hamilton, George-Levi Gayle, Robert Miller, and Limor Golan for their helpful comments and invaluable advice. I have also received helpful comments from Mariagiovanna Baccara, Brian Bergfeld, Chaoran Chen, Jane Cooley Fruehwirth, Ignacio Esponda, Daniel Gottlieb, Robert Munk, Juan Pantano, Nicholas Papageorge, Stephen Ryan and David Wiczer. I have also benefited from presentations at Washington University in St. Louis, University of Michigan, University of North Carolina–Chapel Hill, University of Queensland, Banco de la República, Cornerstone Research, the Centre for European Economic Research, the 2016 EconCon at Princeton, and the 2017 Summer Meeting of the Econometric Society. I also thank the Center for Research in Economics and Strategy at the Olin Business School for financial support during part of this research.

# 1 Introduction

Less than 35% of individuals start a business during their careers and they tend to do so well into their thirties (Table 1 below). Although our understanding of the relative importance of the economic forces explaining these facts is limited, policy makers seek to harvest potential economic gains from entrepreneurship, for long considered an engine of innovation and growth (Schumpeter, 1911), often by focusing on the young.<sup>2</sup> Without a proper understanding of the mechanisms behind the stylized facts it is unclear to what extent entrepreneurship policies induce young people to start businesses, what types of entrepreneurs they attract (e.g., high or low ability), and what the long term consequences of these policies are. Ultimately, it is unclear why the policies should target young individuals.

I address these questions using a life-cycle model of occupational choice. In the model risk-averse individuals sequentially choosing paid employment or entrepreneurship face two types of uncertainty: irreducible idiosyncratic uncertainty, and reducible uncertainty about their own ability. Returns to experience, non-pecuniary benefits, entry costs, and the ability to smooth consumption also determine occupational choices in the model. I provide two main contributions to the literature. First, I quantify the relative importance of determinants of entrepreneurial participation and the gap in first-entry ages between entrepreneurship and paid employment. Importantly, I include risk aversion, which has been overlooked in empirical, dynamic models of entrepreneurship but plays a key role preventing participation. Second, I provide a framework to undertake ex-ante evaluations of the short and long term effects of entrepreneurship policies, and show that fostering young entrepreneurship can offer higher returns. In addition to these contributions, I provide a measure for how much individuals can learn about their entrepreneurial ability before becoming entrepreneurs. To the best of my knowledge, I am the first to provide such measure.

A number of economic forces have been suggested in the literature to explain why individuals,

---

<sup>2</sup> For instance, the OECD recommends that training approaches to foster young entrepreneurship should include experimentation (OECD, 2013). The popular press has often shared the interest in young entrepreneurship (for example, in the *30 under 30* collection by *Forbes*).

both young and old, attempt an occupation, and in particular entrepreneurship. Learning-by-doing, commonly characterized as experience accumulation that increases productivity, is one such force (Keane and Wolpin, 1997; Lazear, 2005; Lafontaine and Shaw, 2016). Learning about one's entrepreneurial ability also determines entry (Jovanovic, 1979; Antonovics and Golan, 2012). 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 because they want to learn whether they are in the "right" part of the distribution. The option value of entrepreneurship, which intersects both learning-by-doing and learning about ability, also affects entry. Individuals attempt entrepreneurship because they can switch back to paid employment if they discover that entrepreneurship is not the best option for them (Manso, 2016), although they are reluctant to experiment if failure is penalized. Risk aversion pushes individuals away from entrepreneurship, which is a more uncertain occupation (Iyigun and Owen, 1998; Hall and Woodward, 2010), and affects the dynamic value of entrepreneurship. 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; Buera, 2009). Finally, non-pecuniary motivations such as "being one's own boss" could also provide incentives for entry (Hamilton, 2000; Hurst and Pugsley, 2015).

Some of the forces explaining entry have predictions that are at odds with the stylized fact that individuals attempt entrepreneurship for the first time in their mid thirties, after accumulating several years of paid employment experience. In the absence of transferability of skills, learning-by-doing implies that individuals who want to become highly productive entrepreneurs should start at an early age. Learning about ability suggests that high ability variance in entrepreneurship encourages individuals to seek to discover their place in the distribution as early as possible (Miller, 1984). However, both of these predictions are made in isolation from other mechanisms such as risk aversion and credit constraints. For instance, risk aversion, which prevents entry at any stage in an individual's career, can have an attenuated effect as individuals acquire more experience if

learning about ability reduces uncertainty over time.<sup>3</sup> Moreover, credit constraints, which preclude individuals with weaker credit histories or lower disposable wealth from entering, may particularly hurt young individuals.

Using data for white and black men between the years 1968 and 1996 from the Panel Study of Income Dynamics (PSID), I estimate a structural model that separates the mechanisms mentioned above.<sup>4</sup> I disaggregate paid-employment occupations into blue collar and white collar, and entrepreneurship into incorporated and unincorporated to capture differences in the abilities required in these two types of entrepreneurship (Levine and Rubinstein, 2017). In estimation, because individuals select based on beliefs as opposed to ability, panel data and occupation-specific first-differences estimators cannot correct for selection bias.<sup>5</sup> I endogenize selection and use the likelihood function implied by the model. I decrease the computational burden from this approach by following a two-stage method that combines an Expectation-Maximization (EM) algorithm and a conditional choice probabilities (ccp) estimator (Arcidiacono and Miller, 2011; James, 2011). The EM algorithm in the first stage bypasses the need for multidimensional integration over unobserved ability vectors. The ccp estimator in the second stage adds flexibility to the treatment of the state space because the structural parameters can be estimated without solving the dynamic optimization problem at every candidate parameter vector during the search algorithm (Hotz and Miller, 1993). In addition, I suggest a novel use for the representation of the dynamic problem in terms of ccps that goes beyond estimation: I employ the representation in counterfactuals that change the regime (as opposed to one-period, unexpected changes, often used for simplicity).<sup>6</sup>

---

<sup>3</sup> Consider the case where paid-employment and entrepreneurial ability are positively correlated. In this case, favorable paid-employment outcomes may be associated with switching into self-employment. Alternatively, if paid-employment outcomes are uninformative of one's entrepreneurial ability, successful workers, already inclined to stay based on their success, have even less incentives to switch because entrepreneurship becomes more risky relative to paid employment.

<sup>4</sup> I abstract from other mechanisms such as personality traits (Hamilton et al., 2018; Humphries, 2018) and parental influence (Lindquist et al., 2015).

<sup>5</sup> Since individual's beliefs change over time as they acquire information, the unobserved component cannot be controlled for using fixed effects (Gibbons et al., 2005).

<sup>6</sup> This implies that the representation must be updated with ccps that capture the nature of the new regime.

Estimates indicate that entrepreneurial ability displays higher variation than paid-employment ability, which both encourages individuals to attempt entrepreneurship in order to learn whether they are high-ability, and discourages them on the basis of risk aversion. However, this ambiguity of incentives can be resolved over time because individuals can use their white collar success as an indicator of entrepreneurial ability before becoming entrepreneurs. For instance, individuals with more than college education can reduce initial uncertainty about incorporated entrepreneurial ability by about 30% after 5 years of white collar experience.

The deterring effect of risk aversion is non negligible. Using the structure of the model, a decomposition exercise indicates that disabling the effect of risk aversion increases the share of individuals who attempt incorporated entrepreneurship by 40%. By comparison, providing full information about ability increases the share by 35%. In fact, if individuals were myopic, i.e. if they discounted the future completely, their certainty equivalent for entrepreneurship would be negative, which is much lower than the positive certainty equivalent of forward-looking individuals. This result underscores the mitigating power of dynamic considerations about future human capital and information. Other factors have even stronger effects on the participation margin such as learning-by-doing and entry costs.

The decomposition also reveals that the main forces explaining the gap in first entry ages between entrepreneurship and paid employment are entry costs and information frictions. The former is consistent with the literature and the latter is a novel result. I find that providing full information about ability induces individuals to enter entrepreneurship earlier, closing the first-entry age gap between white collar work and entrepreneurship by 20%. Entry costs play a strong deterring role and they capture barriers to entrepreneurship not explicitly modeled, such as credit constraints. In order to make the connection between entry costs and credit constraints, these costs are interacted with age and a permanent wealth component (estimated as a fixed effect outside of the structural model). Younger individuals as well as individuals with lower permanent wealth face higher barriers to entry. Flattening the profile of entry costs, which effectively causes individuals of all ages to face the same average entry cost, closes the gap in average first-entry age by about 70%.

Building upon the decomposition results, I study two counterfactual policies aimed to foster young entrepreneurship, thereby closing the gap in first entry ages.<sup>7</sup> In addition, I use the counterfactual policies to analyze the motivation behind promoting young entrepreneurship. The first counterfactual introduces a blanket subsidy for young incorporated entrepreneurship that increases participation and has a small positive effect on the average present value of income (PVI) net of the subsidy. The second counterfactual extends a strand of literature that shows that entrepreneurship education, as a source of information, shifts individual beliefs (von Graevenitz et al., 2010; Oosterbeek et al., 2010). I provide a mapping from movements in beliefs, generated by entrepreneurship education of a given quality, into career choices and long-term outcomes. I find that entrepreneurship education that provides information about incorporated entrepreneurial ability can generate sizable increases in young incorporated entrepreneurship and PVI. Since the information quality of any specific policy may be different and its cost may well exceed the additional income it generates, I use the structure of the model to calibrate the information quality of the entrepreneurship program studied by von Graevenitz et al. (2010). I find that, offered to all students, such a program would increase young incorporated entrepreneurship from 2% to approximately 10%, and it would add \$74,000 to the PVI of the average individual. I use the second counterfactual to show that the gains from entrepreneurship education programs decrease rapidly with age as measured by the PVI. An intervention at age 40 only generates 17% of the \$74,000 gain obtained from an early-career intervention. This result shows that promoting young entrepreneurship is at least partially motivated by the information value of early signals.

In the literature, my paper is closest to Dillon and Stanton (2017) who independently study the choices between entrepreneurship and paid-employment using a learning framework and the PSID. There are four major differences from their work. First, I introduce risk aversion in the model, separating it from specific preferences for entrepreneurship.<sup>8</sup> I find that risk aversion plays

---

<sup>7</sup> Young entrepreneurs are defined as those who attempt entrepreneurship during the first five years of their labor market careers.

<sup>8</sup> I also introduce a consumption smoothing technology that interacts with risk aversion preferences. In the absence of consumption smoothing, a lower risk aversion parameter would be needed to explain low participation rates,

an important role preventing participation in entrepreneurship. Second, I allow for learning to occur in both paid-employment and entrepreneurial occupations, much in the spirit of Jovanovic (1979) and Miller (1984), avoiding asymmetric assumptions about initial information—regarding occupational abilities—that are difficult to justify empirically.<sup>9</sup> Third, I fully control for selection on beliefs in estimation. I do this by using the selection structure of the model, which allows me to obtain consistent estimates without having to impose null cross-occupation effects of experience.<sup>10</sup> Finally, although I allow preferences for entrepreneurship to vary flexibly with observables, I do not allow for unobserved heterogeneity in preferences. For these reasons I see our contributions as complementary.

The rest of the paper is organized as follows. Section 2 presents the data and describes the main regularities motivating the research questions and modeling choices. Section 3 describes the model and its implications. Section 4 discusses identification and describes the estimation method. Section 5 presents the estimated parameters and Section 6 introduces the decomposition exercise, focusing on the mechanisms behind the low participation in entrepreneurship and the gap in first entry ages. Section 7 studies the effects of policies that foster young entrepreneurship, thereby closing the gap, and shows how these policies can offer larger gains. Section 8 concludes.

## 2 Data

The empirical analysis relies on a sample of yearly data for white and black men between the years 1968 and 1996 from the Panel Study of Income Dynamics (PSID). The survey includes information on occupation, self-employment status, business ownership, incorporation status, income, working hours and (in some years) wealth. After dropping observations of individuals who lack data on relevant variables, the sample contains 1,506 unique individuals and 21,334 individual-year observations provided that entrepreneurial outcomes are more variable.

<sup>9</sup> I avoid assuming individual certainty over a subset of one's occupational ability. I do not impose information spillovers across occupations, instead I let the data identify the degree of information spillovers.

<sup>10</sup> Since experience is endogenous to unobserved beliefs, setting the experience coefficient to zero in an income equation assumes away selection concerns. I find strong evidence against this assumption.

vations. About 22% of individuals in the final sample are African American, 42% have college education or more, and on average they enter the labor market at age 22.<sup>11</sup>

Using the data on occupation, self-employment status, business ownership, and incorporation status, both salaried employment and entrepreneurship are disaggregated further to exploit differences in returns to experience as well as differences regarding the information each occupation provides about individual ability. For paid employment, three-digit occupation codes are aggregated into blue collar and white collar (Keane and Wolpin, 1997). For self-employment, which is interchangeably referred to as entrepreneurship in this study, individuals are split between incorporated and unincorporated (Levine and Rubinstein, 2017).<sup>12</sup> Income for the paid employed corresponds to their reported annual labor income. For the self-employed, measuring income is less transparent. Since incorporated individuals are not asked about their business income in the survey, I use their reported labor income, which corresponds to what Hamilton (2000) terms “the draw” or the difference between net profit and retained earnings. For unincorporated individuals, who are not sheltered from the losses of their ventures through limited liability, income corresponds to the sum of labor and business income, and can be zero or negative. Hourly income is obtained by dividing annual income by annual working hours. Wealth data in the PSID is collected only on a few periods before 1996. Thus, I construct a measure of permanent wealth, denoted  $\omega$ , using all wealth observations available (including the ones after 1996), in order to capture long-run differences in access to resources. This measure is defined as the constant plus the fixed effect of a regression of wealth on a second degree polynomial on age. Median (mean) permanent wealth is

---

<sup>11</sup> The PSID turned biennial after 1996. To avoid assumptions about occupational choices and income in years where no data were collected, I do not employ data after 1996. Notably, dropping observations of individuals who are not observed early in their careers renders the sample not representative in relation to the original PSID sample (Appendix A). In the learning framework of this paper, modeling missing yearly data as unobserved would imply cumbersome integration over long sequences of unobserved signals.

<sup>12</sup> In the sample, 79 (87) percent of unincorporated (incorporated) individuals respond affirmatively to the business ownership question in the PSID, 41 (22) percent belong to construction and repair industries, and 7 (22) belong to financial services, real estate and manufacturing. Corporations are separate legal entities from their owners, offer limited liability, and can raise money by issuing stocks. However, the advantages of incorporation come at the costs of more complex administrative activities and higher administrative costs.

300 (400) thousands of dollars.<sup>13</sup>

**TABLE 1: 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:  $exp_k$  stands for experience in occupation  $k$ ; for instance,  $exp_{ei}$  stands for incorporated entrepreneurial experience. 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.

Entrepreneurship is much less common in an individual's career than paid employment (Table 1). The proportion of individuals who attempt entrepreneurial occupations is less than half the proportion of individuals who attempt salaried occupations.<sup>14</sup> Furthermore, most of the 4,294 occupational spells in the sample are paid-employment spells, which are more than 60 percent longer than entrepreneurial spells (Table 2). Table 1 also illustrates the lack of young entrepreneurship in the data. Those who attempt entrepreneurial occupations do so later in their careers, after accumulating more than 8 years of paid-employment experience on average. This opens a gap in average entry age between entrepreneurial and salaried occupations that runs opposite to the prediction of a parsimonious model of uncorrelated learning about ability (Miller, 1984). Notably, although Table 2 indicates that few individuals start their careers as entrepreneurs, Figure 1 shows that participation increases as individuals age.

Different types of entrepreneurs vary on their patterns of transition into paid employment. Table 3 shows that unincorporated entrepreneurs are just as likely to switch to either blue collar or white collar work, while incorporated entrepreneurs tend to transition to white collar work. Consistent with the higher complexity of the incorporated organizational structure, Table 4 shows that

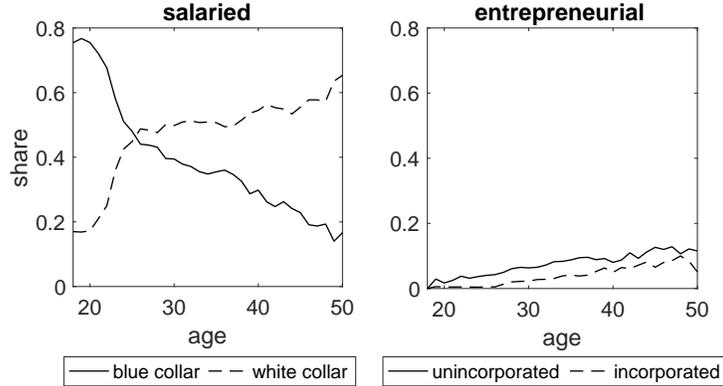
<sup>13</sup> This measure can be thought of as a complicated function of initial wealth, understood broadly, and the environment, and it is akin to the fixed effect in the equation of optimal hours in a lifecycle model of labor supply in MaCurdy (1981). The level of the permanent wealth measure depends on the degree of polynomial implemented. The estimated value of the constant is about \$418,000 (Appendix A).

<sup>14</sup> 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 by then.

**TABLE 2: 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.

**FIGURE 1: Occupational Choice: Participation by age**

incorporated individuals, as well as white collar workers, tend to be more educated than unincorporated individuals. Entrepreneurs are more likely to be older, white, married, and work more hours than paid employees, but they have higher hourly income on average. For instance, average hourly income for incorporated entrepreneurs is 75% higher than for white collar workers; this is due in part to the fact that individuals try entrepreneurship later in their careers. Table 4 also shows that entrepreneurship displays higher hourly income variation than paid-employment. Even after controlling for demographics and occupation-specific experience fully interacted with occupation, the variance of hourly income in incorporated entrepreneurship is more than three times as large as the variance in white collar work.

**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  $t + 1$ .

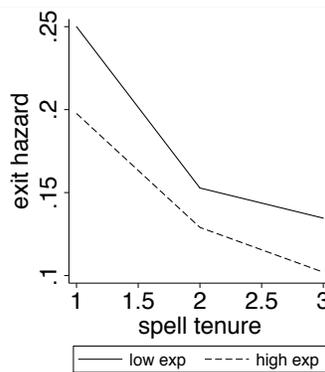
Paid-employment experience is associated with entry, exit, and success in entrepreneurship.

**TABLE 4: Sample Characteristics per Occupation**

	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
Age	31.04 (7.27)	28.92 (6.65)	32.21 (7.18)	33.93 (7.35)	36.94 (7.00)	30.45 (8.06)
Black	0.20	0.31	0.12	0.08	0.07	0.46
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
Hours Worked	2147 (693)	2096 (617)	2234 (559)	2329 (819)	2703 (724)	
Hr Labor Income	18.71 (16.12)	14.16 (7.97)	21.24 (14.26)	21.30 (23.08)	37.91 (51.69)	
Residual Hr Labor Income		(6.99)	(12.32)	(20.41)	(44.29)	

Notes: Standard deviation is in parentheses. Individual-year observations summarized by occupation. Monetary quantities are in real dollars of 2000. 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.

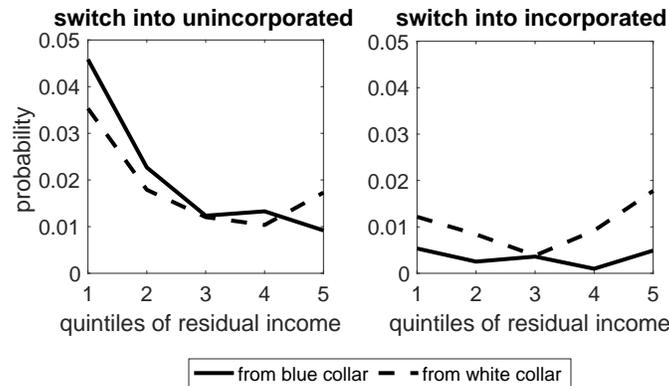
Figure 2 shows that the exit hazard during the initial years of the first entrepreneurship spell is lower for individuals with high (more than six years) prior paid-employment experience. Additionally, it cannot be rejected at five percent significance that those with high prior paid-employment experience have higher (0.2 standard deviations) residual income in their first entrepreneurship spell. These facts are consistent with salaried occupations being informative of entrepreneurial ability as individuals entering entrepreneurship for the first time after acquiring paid-employment experience will be better selected, and therefore will be more successful and exit less frequently.

**FIGURE 2: Exit Hazard from First Spell of Entrepreneurship by Salaried Experience**

Notes: Probability of switching out of first spell of entrepreneurship in  $t + 1$  conditional on surviving until  $t$ , by prior salaried experience. Low (high) corresponds to  $\leq$  ( $>$ ) 6 years of salaried experience prior to first spell of entrepreneurship.

At the entry margin, Figure 3 shows two features. First, higher residual income in either white

collar or blue collar work is generally associated with a smaller probability of switching into unincorporated entrepreneurship.<sup>15</sup> Second, among white-collar workers, those with the highest residual income are the most likely to switch into incorporated entrepreneurship. Although the first feature is consistent with uncorrelated learning where unexplained success is only informative of ability in the current occupation, the second feature and results in the previous paragraph are consistent with positive correlated learning between white collar work and incorporated entrepreneurship. This highlights the importance of allowing the data determine the degree of correlated learning.



**FIGURE 3: 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.

In the remaining of the paper I built and estimate a model that explains the mechanisms behind two main stylized facts presented in this section: most people do not attempt entrepreneurship and there is a gap in first entry ages between entrepreneurship and paid employment. Other stylized facts in this section motivate the model selection and exemplify the sources of identifying variation in the data.

<sup>15</sup> Residual income obtained from occupation-specific OLS regressions after controlling for demographics and a quadratic in occupation-specific experience.

### 3 Model

In the model, forward-looking individuals face dynamic incentives that reflect accumulation of experience and information. 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. Individuals are risk averse and face irreducible idiosyncratic uncertainty as well as uncertainty about their own ability, which they can reduce by observing their labor market outcomes. Their productivity in each occupation is determined by their experience, their unobserved ability, and idiosyncratic shocks. Hence, they receive noisy income signals with which they update beliefs about their own ability. Individuals maximize expected utility, which they compute using their beliefs, and are able to smooth consumption over time.

#### 3.1 Occupations, Income, and Individual Characteristics

An individual in the model, denoted  $i$ , enters the labor market at age  $t_{i0}$  and at every age  $t = t_{i0}, \dots, T$  chooses an occupation  $k \in \{0, 1, \dots, K\}$ , where  $k = 0$  denotes the choice of not to work. Occupations can be salaried or entrepreneurial. Let  $d_{kit} \in \{0, 1\}$  be an indicator for whether or not he chooses occupation  $k$  at age  $t$ . In the remaining exposition of the model I drop the individual indicator  $i$ . The individual accumulates experience, collected in a  $K$ -dimensional vector denoted  $x_t$ , that partly determines his productivity in each occupation. He starts his career with no experience and from that point forward his experience in occupation  $k$  evolves as a function of his choices:

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

When an individual does not work he receives no income. Alternatively, when he works in occupation  $k > 0$ , he receives income  $\bar{L}_k y_{kt+1}$  at the beginning of next period, where  $\bar{L}_k$  denotes the

number of annual hours that all individuals supply inelastically if they decide to work in occupation  $k$ , and  $y_{kt+1}$  denotes the individual-specific hourly income.<sup>16</sup> His hourly income in occupation  $k > 0$  is not constrained to be positive (for instance, if an entrepreneur has losses), and is the sum of three components, a mapping  $f_k(\cdot) : \mathbb{Z}_{\geq 0}^K \rightarrow \mathbb{R}$  of his experience vector in addition to his fixed, occupation-specific ability  $\mu_k \in \mathcal{M}$ , and a temporary productivity shock denoted  $\eta_{kt+1}$ :

$$y_{kt+1} = f_k(x_t) + \mu_k + \eta_{kt+1} \quad (2)$$

His experience in other occupations  $k' \neq k$ , i.e. his *cross-occupation experience*, affects his productivity in occupation  $k$  if  $\frac{\partial f_k(\cdot)}{\partial x_{k'}} \neq 0$ . His occupation-specific abilities are collected in a  $K$ -dimensional vector  $\mathcal{M}$  drawn from the population distribution of occupational ability characterized by the cumulative distribution function  $F_{\mathcal{M}} : \mathbb{R}^K \rightarrow [0, 1]$ . Productivity shocks  $\eta_{kt+1}$  are drawn independently across individuals, periods, and occupations, from an occupation-specific distribution characterized by the cumulative distribution function  $F_{\eta_k} : \mathbb{R} \rightarrow [0, 1]$ .

### 3.2 Information Structure and Learning

Both the distribution of occupational ability and the distributions of productivity shocks are common knowledge. Let  $\mathcal{I}$  denote the amount of additional information that the individual has about his occupational ability at the beginning of his career, and assume that he has rational expectations. Then, his beliefs at the beginning of his career are given by  $\mathbb{B}_{t_0} = F_{\mathcal{M}|\mathcal{I}}$ . For instance, if  $\mathcal{I} = \{\mu_{k'}\}$  for some  $0 < k' \leq K$ , then he knows his ability in one of the occupations. Alternatively, if  $\mathcal{I} = \emptyset$  then he only knows the population distribution of occupational ability and must learn about each of his occupational abilities by participating in the labor market.<sup>17</sup>

The individual only observes a temporary, occupation-specific productivity shock after joining

<sup>16</sup> In this sense the model abstracts from the working hours decision given an occupation.

<sup>17</sup>  $\mathcal{I}$  can also denote cases in which the individual has received information signals before entering the job market. Moreover, if  $\mathcal{I} = \{\mathcal{M}\}$  then there is full information about occupational ability. Lafontaine and Shaw (2016) use this assumption when studying entrepreneurship using fixed-effects effects regressions.

an occupation.<sup>18</sup> Furthermore, he does not separately observe his occupational ability from his temporary shock, which prevents him from learning his ability immediately after working for one period. In other words, he cannot distinguish how much of his unexplained productivity comes from his occupational ability and how much comes from luck. Formally, he observes a signal  $\zeta_{kt+1} \equiv y_{kt+1} - f_k(x_t)$ , after choosing occupation  $k$ . Then, given his prior beliefs  $\mathbb{B}_t$ , he updates his beliefs using Bayes' Rule as follows

$$d\mathbb{B}_{t+1}(\mathcal{M}) \equiv f_{\mathcal{M}|\zeta_{kt+1}, \mathbb{B}_t}(\mathcal{M}|\zeta_{kt+1}, \mathbb{B}_t) = \frac{f_{\zeta_{kt+1}|\mathcal{M}, \mathbb{B}_t}(\zeta_{kt+1}|\mathcal{M}, \mathbb{B}_t) d\mathbb{B}_t(\mathcal{M})}{\int_{\mathcal{M}'} f_{\zeta_{kt+1}|\mathcal{M}', \mathbb{B}_t}(\zeta_{kt+1}|\mathcal{M}', \mathbb{B}_t) d\mathbb{B}_t(\mathcal{M}')} \quad (3)$$

where  $d\mathbb{B}_{t+1}$  stands for the probability density function associated with his beliefs at  $t + 1$ , and  $f_{\zeta_{kt+1}|\mathcal{M}, \mathbb{B}_t}$  is the probability density function of the signal received from occupation  $k$  at  $t + 1$ , conditional on a value for  $\mathcal{M}$  and his beliefs at  $t$ .<sup>19</sup>

The information framework summarized by initial prior  $\mathbb{B}_{t_0} = F_{\mathcal{M}|\mathcal{I}}$  and the updating rule in equation (3) captures as special cases learning about occupational abilities that are unobserved and uncorrelated (Miller, 1984), unobserved and correlated (James, 2011), and partially observed and correlated (Dillon and Stanton, 2017). This general framework allows for information spillovers to occur gradually during the life-cycle, depending on the assumptions on  $\mathcal{I}$  and  $F_{\mathcal{M}}$ . For instance, an individual may use the information received as a salaried worker to update his beliefs about his entrepreneurial ability, I denote this type of learning *cross-occupation learning about ability*.<sup>20</sup>

<sup>18</sup> Hence, equation (2) implies that 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.

<sup>19</sup> Given the definition of  $\zeta_{kt+1}$ ,  $f_{\zeta_{kt+1}|\mathcal{M}, \mathbb{B}_t}$  also depends on the distribution of the productivity shock. Specific updating rules for the problem can be found in Appendix B.

<sup>20</sup> Notice that information spillovers prevent the researcher from representing the occupation choice problem in terms of dynamic allocation indexes à la Miller (1984).

### 3.3 Utility

The individual is forward-looking and his discount factor for future utility is  $\beta$ . His flow utility  $u(\cdot, \cdot; \rho) : \mathbb{R} \times \mathbb{R}^{K+1} \rightarrow \mathbb{R}$  depends on consumption  $c_t$  and a vector of observed, alternative-specific utility shocks  $\varepsilon_t$  drawn from a distribution characterized by the cumulative distribution function  $F_\varepsilon : \mathbb{R}^{K+1} \rightarrow [0, 1]$ , which is known. His utility function is characterized by the risk aversion parameter  $\rho$ . Additionally, the support of  $u(\cdot, \cdot; \rho)$  allows for the individual to have negative consumption. This may occur when his income is negative (for instance, if his business losses money) and he has no consumption smoothing technology available.

The utility function incorporates the fact that the individual may be less willing to try occupations that are more risky, as often entrepreneurial occupations are thought to be. Risk aversion interacts with the information structure in interesting ways. On the one hand, the individual always dislikes believed future income variance that is due to temporary productivity shocks. On the other, when his believed income variance is due to occupational ability, he faces a tradeoff between information value and risk aversion because, before knowing his place in the distribution, he likes occupations where there is high variance in occupational ability. Moreover, if there is cross-occupation learning about ability, he may choose to optimally sample other occupations before the one with the highest variance to avoid harsher utility costs from risk aversion while reducing the uncertainty around his beliefs. This path can become even more attractive if the effects of cross-occupation experience are positive.

### 3.4 Specification

This section introduces the set of modeling assumptions that further specify the model. The individual has a vector of observed (to the econometrician) characteristics  $h_t$  containing his race (white or black), education level (high school or less, some college, college, and more than college), age, marital status (over which he has perfect foresight), and accumulated experience. The distribution of preference shocks  $F_\varepsilon$  is Type I Extreme Value with location and scale parameters normalized

to 0 and 1, respectively; preference shocks are independent across alternatives, individuals, and over time. Occupational ability is distributed  $N(\mathbf{0}, \Delta_s)$ , allowed to vary by education level  $s$  since individuals with different education levels are likely selected based on ability. Occupation-specific productivity shocks are distributed  $N(0, \sigma_{\eta_k}^2)$ . These normality assumptions in levels (as opposed to log normality) allow for entrepreneurial losses. Individuals start their careers with no additional information about their ability. Hence, by virtue of the joint normality of the distribution of occupational ability, his initial beliefs can be characterized by the mean and variance of the population distribution:  $\mathbb{B}_{t_0} = \langle \mathbf{0}, \Delta_s \rangle$ . Moreover, the normality of both  $F_{\mathcal{M}}$  and  $F_{\eta_k}$  yields a posterior which is also multivariate normal. Therefore, at any age  $t$  his beliefs are summarized by the posterior mean and variance.

If he decides to work, the individual can be a white collar or blue collar paid employee, or he can be unincorporated or incorporated self-employed—hence,  $K = 4$ . The hourly income mapping  $f_k(\cdot)$  is linear in a vector of parameters  $\theta_k$ , is augmented to include all the components of  $h_t$ , and captures the profile of returns to experience using step-functions of the experience vector  $x_t$ . The individual is infinitely lived. He works until age  $T$  and begins his retirement at  $T + 1$ .<sup>21</sup> His flow utility  $u(\cdot, \cdot; \rho)$  is characterized by a CARA function of consumption  $c_t$  where  $\rho$  is the absolute risk aversion.<sup>22</sup> The flow utility mapping is augmented to include  $h_t$ , and the marginal contribution of consumption to his utility is occupation-specific and determined by the non-pecuniary cost of each occupation,  $\alpha_{kt}$ . Concretely, the individual’s flow utility from alternative  $k$  is

$$-\alpha_{kt}(h_t) \exp\{-\rho c_t - \varepsilon_{kt}\} \quad (4)$$

where

$$\alpha_{kt}(h_t) = \exp\{\alpha_{k0} + \alpha_{k1}black + \alpha_{k2}married_t + 1\{x_{kt} = 0\}(\alpha_{k3} + \alpha_{k4}t + \alpha_{k5}\omega + \alpha_{k6}t\omega)\} \quad (5)$$

Although for notational simplicity I write  $\alpha_{kt}(h_t)$ , the non-pecuniary component includes a first-time entry cost that depends on permanent wealth ( $\omega$ ) and age. This is a reduced-form way of

<sup>21</sup> Retirement age is set at  $T + 1 = 51$  for data availability reasons.

<sup>22</sup> Individuals in reality may become less risk averse once they amass enough wealth. Given that the model is already fairly complex I abstract from this mechanism using a CARA specification, which yields substantial gains in tractability (Proposition 1).

capturing entry barriers that are not explicitly modeled. The age profile in the entry cost is meant to capture short-run difficulties that young individuals with weaker credit histories and less savings often face when attempting entrepreneurship. Additionally, long-run advantages in access to resources reducing entry costs to entrepreneurship are captured with the permanent wealth measure.

The consumption smoothing set up follows Margiotta and Miller (2000) relaxing, in a tractable fashion, the hand-to-mouth assumption used in other life-cycle models of entrepreneurship (Dillon and Stanton, 2017; Humphries, 2018). Individuals in the model have access to a contingent-claims market for consumption goods to smooth their consumption using their wealth.<sup>23</sup> Let  $\lambda_\tau$  denote the derivative of the price measure for claims to consumption at date  $\tau$ .<sup>24</sup> Conditional on the individual's choice at period  $t$ ,  $d_t$ , the budget constraint in equation (6) provides the law of motion for disposable wealth  $\xi_{t+1}$ , reflecting how the individual allocates his financial resources between current period consumption and next period savings.

$$E_t[\lambda_{t+1}\xi_{t+1}|d_t, h_t, \mathbb{B}_t] + \lambda_t c_t \leq \lambda_t \xi_t + E_t[\lambda_{t+1}\bar{L}y_{t+1}|d_t, h_t, \mathbb{B}_t] \quad (6)$$

This consumption smoothing setting captures in a more realistic fashion the consumption choices of individuals and the determinants of occupational choices, preventing the econometrician from understating the effects of risk aversion. For instance, if entrepreneurship is the most risky occupation, a hand-to-mouth model would explain the lack of participation with a lower level of risk aversion since individuals would have to absorb every period the entire variation in income.

---

<sup>23</sup> Following Margiotta and Miller (2000), income is assumed uninsurable to capture unobservable insurance risk. This assumption could also capture unobserved levels of effort in labor supply (Green, 1987).

<sup>24</sup> The commodity space for consumption goods is formed by consumption units at date 0 and claims to consumption at calendar date  $\tau$  contingent on how history unfolds.  $\lambda_\tau$  denotes the derivative of the price measure for claims to consumption at date  $\tau$ ,  $\Lambda_\tau$ . Therefore, the price of a unit of consumption to be delivered with certainty at date  $\tau$  in terms of date 0 consumption units is  $E[\lambda_\tau]$ .  $\bar{L}_k$  is specified as the average number of hours worked by individuals in occupation  $k$  in the sample.

### 3.5 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  $\varepsilon_t$ . 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.<sup>25</sup> Similar to results in Margiotta and Miller (2000) and Gayle et al. (2015), the consumption smoothing setting 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 (Proposition 1). This expression satisfies a portfolio separation property according to which individuals hold only a few securities: a bond  $b_\tau$  and a security  $a_\tau$  that pays a random quantity.<sup>26</sup> 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  $t$ . Proposition 1 provides the ex-ante value function of an individual at any age before retirement, when he is choosing an occupation in addition to his consumption and asset portfolio.<sup>27</sup>

**Proposition 1.** *At any age  $t$  before retirement,  $t \leq T$ , the value function of an individual who has*

<sup>25</sup> Upon retirement, he simply decides on his consumption and assets portfolio in order to smooth his remaining wealth.

<sup>26</sup> See, for instance, Altuğ and Labadie (1994).  $b_\tau$  is a bond that, contingent on the history through calendar date  $\tau$ , pays a unit of consumption from period  $\tau$  in perpetuity in date- $\tau$  prices.  $a_\tau$  is a security that pays the random quantity  $(\ln \lambda_s - s \ln \beta)$  of consumption units from period  $\tau$  in perpetuity, in date- $\tau$  prices. The prices of these assets are given by

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

<sup>27</sup> The ex-ante value function is defined as the value function before knowing the realization of the vector of taste shocks,  $\varepsilon_{it}$ .

not yet observed his taste shocks,  $\varepsilon_t$ , can be written as

$$V_t(h_t, \mathbb{B}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}) = -\lambda_{\tau(t)} b_{\tau(t)} \exp\left(\frac{-(\rho \xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right) A_t(h_t, \mathbb{B}_t) \quad (7)$$

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

$$A_t(h_t, \mathbb{B}_t) = \sum_{k=0}^4 p_{kt}(h_t, \mathbb{B}_t) \alpha_{kt}(h_t)^{1/b_{\tau(t)}} E_{\varepsilon} [e^{-\varepsilon_{kt}^*/b_{\tau(t)}}] E_t [A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1}) v_{kt+1} | \mathbb{B}_t, h_t]^{1-1/b_{\tau(t)}} \quad (8)$$

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

The probability of choosing  $k$  at age  $t$  conditional on characteristics and beliefs is denoted  $p_{kt}(h_t, \mathbb{B}_t)$ .  $\varepsilon_{kt}^*$  is the value of the taste shock  $\varepsilon_{kt}$  conditional on option  $k$  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  $\mathbb{B}_{kt+1}$ ; both are conditional on choosing  $k$  at  $t$ . Proof: See Appendix B.1.

The index  $A_t(h_t, \mathbb{B}_t)$  in Proposition 1 is a strictly positive average of expected outcomes weighted by the conditional choice probability of each option. The function  $v_{kt+1}$ , in the recursive formulation of  $A_t(h_t, \mathbb{B}_t)$ , is a utility measure of income that adjusts for consumption smoothing, capturing the effect of the consumption smoothing technology on occupational choices. Higher values of the prior mean or higher values of human capital are associated with lower values of the index via the term  $v_{kt+1}$ , for a given age  $t$ . The index also varies with age since younger and older individuals differ in the value they allocate to information and human capital. For young individuals the value of human capital or tight beliefs is much higher because they have a longer time ahead of them to exploit their productivity. Finally, the value of human capital and beliefs in occupation  $k$  decreases with the size of the non-pecuniary costs,  $\alpha_{kt}$ .

Using equations (7) and (8), and applying logs to transform the problem, it can be shown that

at any age  $t$  before retirement the individual chooses an occupation  $k'$  to solve

$$\max_{k' \in \{0, 1, \dots, K\}} \sum_{k=0}^K d_{kt} \{ \varepsilon_{kt} - \ln \alpha_{kt}(h_t) - (b_{\tau(t)} - 1) \ln E_t [A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1}) v_{kt+1} | \mathbb{B}_t, h_t] \} \quad (9)$$

His occupational choice, which already accounts for consumption smoothing, 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.<sup>28</sup> Trade-offs between occupations are characterized by differences in non-pecuniary utility components,  $\varepsilon_{kt}$  and  $\alpha_{kt}(h_t)$ , and differences in the expected utility from income,  $v_{kt+1}$ , scaled by the index capturing the value of human capital accumulation and beliefs evolution,  $A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})$ . Since  $F_\varepsilon$  is assumed Type-I Extreme Value, the expression in equation (9) becomes a standard logit. Following Hotz and Miller (1993), Proposition 2 uses the recursive nature of the index  $A_{t+1}(h_t, \mathbb{B}_t)$  and yields a representation of the logarithm of the odds ratio in terms of future choice probabilities and utility 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*

$$\ln \left( \frac{p_{kt}(h_t, \mathbb{B}_t)}{p_{0t}(h_t, \mathbb{B}_t)} \right) = -\ln \alpha_{kt}(h_t) - (b_{\tau(t)} - 1) \ln E_t \left[ v_{kt+1} \prod_{s=1}^{T-t} \left( \frac{p_{0t+s}(h_{kt}^{(s)}, \mathbb{B}_{kt}^{(s)})}{p_{0t+s}(h_{0t}^{(s)}, \mathbb{B}_{0t}^{(s)})} \right)^{\phi_t(s)} \middle| \mathbb{B}_t, h_t \right] \quad (10)$$

where

$$\phi_t(s) = \frac{1}{b_{\tau(t)+s}} \prod_{r=1}^{s-1} (1 - 1/b_{\tau(t)+r}) \quad (11)$$

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

Equation (10) 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 (9).

<sup>28</sup> Modifying the model to allow for 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.

Higher non-pecuniary costs and lower expected utility from compensation, perhaps due to higher uncertainty, make option  $k$  less likely to be chosen. Moreover, if choosing occupation  $k$  makes the individual less likely to work in the future, thereby reducing the value of his human capital or his information, then occupation  $k$  is also less likely to be chosen today. The results from Proposition 2 are exploited in estimation as the next section explains.

## 4 Identification and Estimation

There are three sets of parameters to be identified: returns to experience and education  $\theta_k$ , parameters of the learning structure including the distribution of occupational ability  $F_{\mathcal{M}}$ , and utility parameters  $\{\rho, \alpha_k\}$ . 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. Identification of each set of parameters is briefly discussed below.

Individuals select on beliefs, which poses a challenge for estimation of the returns to occupation-specific experience in the income equations (Gibbons et al., 2005). I endogenize selection and use the likelihood function implied by the model. Hence, income variation for different levels of occupation-specific experience identifies the returns to experience in each of the occupation-specific income mappings  $f_k(\cdot)$ . In particular, the returns to cross-occupation experience are identified from the variation in income of switchers. Notably, since education indicators enter linearly in the income equations, the mean of the distribution of ability, which is specific to each education level, is not identified and normalized to zero.<sup>29</sup> Identification of the latent distribution of ability  $F_{\mathcal{M}}$  under normality assumptions using income data and occupation choices has been shown in Heckman and Honoré (1990). 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 covariance matrix, determining the degree of correlated learning, are identified from the covariation

---

<sup>29</sup> Alternatively, the education-specific shifters in the income equation can be thought of as education-specific means for the ability distribution.

in residuals of switchers. For instance, variation in the mix of prior salaried experience and residual returns at entry into entrepreneurship (Section 2) help identify both cross-occupation returns and the extent of correlated learning between paid-employment and entrepreneurial occupations. Finally, the variance of the distribution of productivity shocks is identified from the excess variation in residual income per occupation.<sup>30</sup>

Identification of the flow payoffs follows from results in Magnac and Thesmar (2002); Arcidiacono and Miller (2015). Given that the distribution of the choice-specific taste shocks, the subjective discount factor, and the transition function of beliefs in equation (3) 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—in other words,  $\alpha_{0t}(\cdot) \equiv 1$ . Hence, the functional form assumptions regarding the utility function provide over-identifying restrictions. Separate identification of the risk aversion parameter and prior beliefs relies on the panel dimension of the data. This is because observationally equivalent choice patterns could be generated by overconfident low-variance priors and high risk aversion, or under confident high-variance priors and low 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.

Estimation of the structural model entails 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). Because individuals select on beliefs rather than ability, conditional on the history of income signals up to  $t$ , mapped into beliefs  $\mathbb{B}_{it}$ , choices at  $t$  are independent of ability. As a consequence of this, the log likelihood of the data can be separated additively, allowing for the two-stage estimation procedure summarized below.<sup>31</sup> The likelihood of the data and a more detailed description of the estimation procedure are presented in Appendix C. Standard errors are corrected for the two-stage estimation procedure using subsampling estimation

<sup>30</sup> This can be seen more clearly in the updating rules of the estimation algorithm in Appendix C.

<sup>31</sup> Additive separability requires that no measurement error is assumed in the hourly income data. Integration over measurement errors breaks the additive separability property.

over 100 subsamples without replacement.

*First Stage.* An Expectation-Maximization (EM) algorithm yields estimates of the collection of income parameters, denoted  $\Theta$ , and the covariance matrix of the population ability distribution conditional on education level  $s$ , denoted  $\Delta_s$ .<sup>32</sup> The EM algorithm is an iterative method that yields maximum likelihood estimates when a portion of the data is unobserved. In this case, the unobserved part of the data is the individual's ability,  $\mathcal{M}$ . The EM algorithm permits fast estimation of income and ability parameters as it avoids multidimensional integration over unobserved ability vectors.

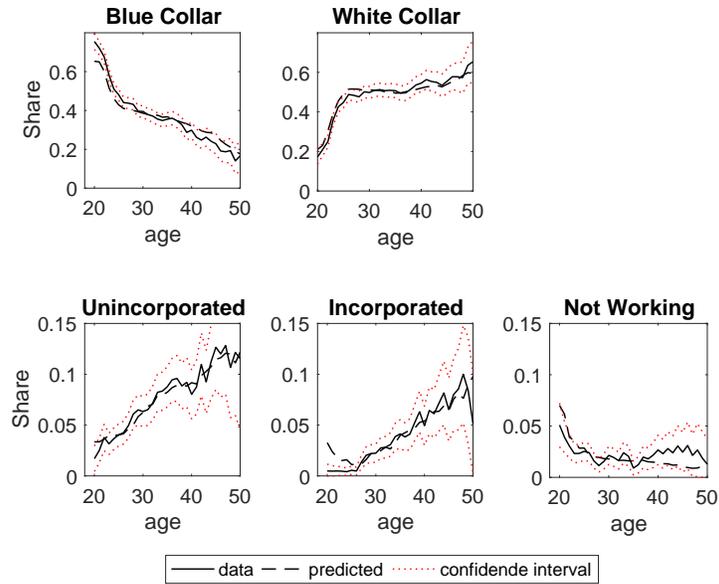
*Second Stage.* A ccp estimator yields estimates of the collection of parameters of the utility function, denoted  $\Lambda$ . The estimator uses Proposition 2 which provides a mapping from future choice probabilities and utility parameters into current choice probabilities. An iterative algorithm is 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).<sup>33</sup> The ccp estimator speeds up estimation and adds flexibility to the treatment of the state space because the structural parameters can be estimated without solving the dynamic optimization problem at every candidate parameter vector during the search algorithm.

The estimated model replicates choices in the data well (Figure 4) with a minor deviation generated by the initial state of very young individuals, below 20 years old. For these individuals, whose completed education is high school or less, the model underestimates participation in blue collar work, in favor of white collar work and incorporated entrepreneurship. This deviation is corrected by age 20. Additional fit statistics in Appendix D show that the model captures well the proportion of individuals who attempt entrepreneurial occupations by age 40, and it captures reasonably well the average age at first entry into all occupations. More interestingly, the model captures the mix of salaried experience obtained before first entry. Although not targeted in estimation, the model

---

<sup>32</sup>  $\Theta$  includes the variance parameters of the productivity shocks,  $\sigma_{\eta_k}^2$ .

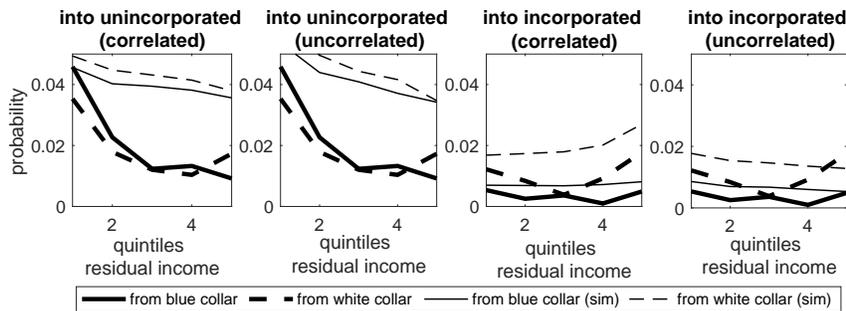
<sup>33</sup> The asymptotic properties of iterative estimators where the first-iteration estimator is consistent (as is the case here) are discussed in Amemiya (1985).



**FIGURE 4: Model Fit**

Notes: Actual and simulated choices by age, taken the state at  $t$  in the data as given for simulation.

is also able to capture the trend—although not the level—in the relation between the probability of switching into entrepreneurship from salaried occupations and the current income signal (Figure 5). The model captures the relative flatness of the relation between the signal and the probability of switching into incorporated entrepreneurship from blue collar work. Additionally, the model captures the increase in the probability of switching into incorporated entrepreneurship from white collar for those receiving the best signals. Figure 5 also shows that a model that does not allow for correlated learning is unable to capture this trend.



**FIGURE 5: Probability of Switching into Entrepreneurial Occupations**

Notes: Probability of switching into entrepreneurial occupations in  $t + 1$  by decile of residual income in  $t$ , from observed and simulated (sim) data. Residual income is computed from occupation-specific OLS regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status. Simulated data generated from the baseline model (correlated learning) and from an alternative counterfactual where learning about ability is uncorrelated.

## 5 Structural Estimates

### 5.1 Returns to Experience and the Distribution of Ability

Learning-by-doing and learning about ability are two key economic forces in the model. Beginning with learning-by-doing, Figure 6a illustrates estimated returns to own-occupation experience.<sup>34</sup> It shows that the returns to blue collar experience are rather flat while the returns to incorporated experience are the steepest. Figure 6a suggest that individuals aspiring to become highly productivity incorporated entrepreneurs must accumulate at least 5 years of incorporated entrepreneurial experience. Additionally, learning-by-doing can happen across-occupations. Figure 6b illustrates estimated returns to cross-occupation experience. It shows that, while expertise in entrepreneurship, as measured by high accumulated experience, increases productivity in paid employment, low levels of entrepreneurial experience can reduce it.<sup>35</sup>

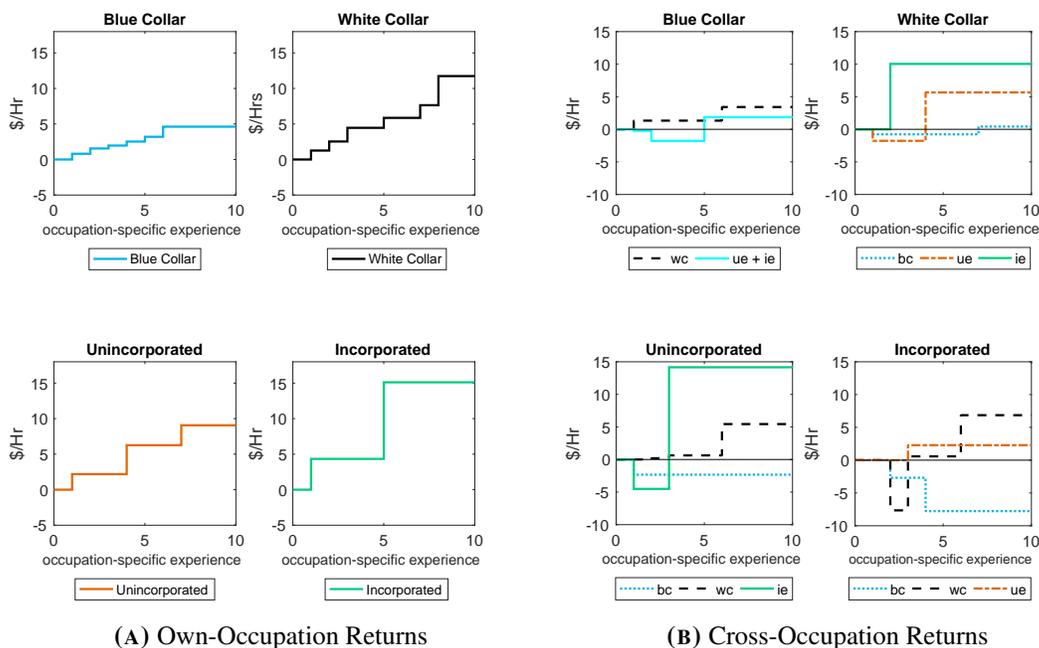
As they work, individuals also learn about their occupation-specific abilities. This learning process is determined by the covariance matrix of their ability distribution. Results suggest that there is more for individuals to learn about in entrepreneurship. In other words, there is higher variation in entrepreneurial ability than in paid-employment ability. Consider the covariance matrix for individuals with more than college education in Table 5.<sup>36</sup> Since ability is measured in \$10s/hr, one standard deviation above the population mean translates into  $10 \times \sqrt{10.88} \approx \$33/\text{hr}$  for incorporated ability and into  $10 \times \sqrt{0.87} \approx \$9/\text{hr}$  for white collar ability. In addition, since more education tends to be associated with higher variation in ability (Table S3 on Appendix D), more educated individuals can potentially benefit more from learning about their entrepreneurial abilities. The off-diagonal terms of the covariance matrix imply that there is correlated learning about ability. Figure 7 shows that individuals with high ability in white collar work tend to have high ability in

---

<sup>34</sup> Estimates for equation (2) are presented in Table S2 in Appendix D

<sup>35</sup> This finding is similar to results in Jovanovic and Nyarko (1996), where switching technologies can reduce productivity by reducing expertise, and is consistent with results in Manso (2016) showing that entrepreneurial experience generates a premium for salaried workers whenever it is above a given threshold.

<sup>36</sup> This result is seen across education levels (Table S3 in Appendix D).



**FIGURE 6: Returns to Experience**

Notes: Returns implied by estimates of equation (2) which is specified using step functions (Table S2 in Appendix D). Occupations in Figure 6b are: white collar (wc), blue collar (bc), unincorporated entrepreneurship (ue), and incorporated entrepreneurship (ie). In blue collar work, experience from both entrepreneurial occupations are pooled. Steps were chosen using statistical significance in a preliminary OLS regression as a baseline. 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.

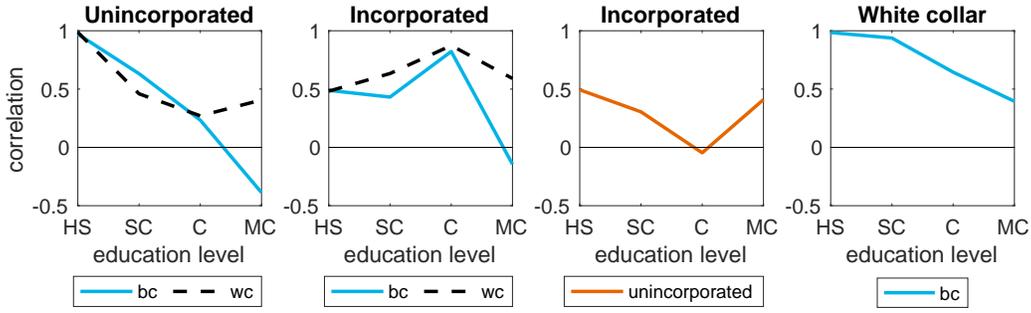
incorporated entrepreneurial activities as well. Interestingly, the correlation between white collar ability and incorporated ability is higher than the correlation between both entrepreneurial abilities at every education level.

**TABLE 5: Population Ability Covariance Matrix (More Than College)**

	More than College							
	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Blue Collar	0.37	(0.044)						
White Collar	0.22	(0.044)	0.87	(0.075)				
Unincorporated	-0.41	(0.128)	0.66	(0.143)	3.03	(0.297)		
Incorporated	-0.29	(0.650)	1.82	(0.125)	2.35	(0.885)	10.88	(1.690)

Notes: Standard errors corrected for 2-stage estimation using subsampling estimation over 100 subsamples without replacement. Covariance matrix of the joint distribution of unobserved ability conditional on education, denoted  $\Delta_s$ .

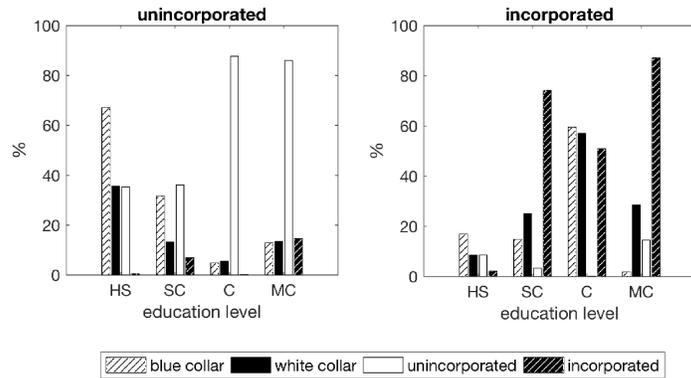
There is an incentive for young individuals at the beginning of their careers to attempt entrepreneurship in order to learn whether they are high-ability. Alternatively, individuals can use their white collar success as an indicator of entrepreneurial ability. However, the informational value of entrepreneurship is undermined by higher idiosyncratic variation in entrepreneurial out-



**FIGURE 7: Implied Correlation Between Abilities**

Notes: Correlation between abilities per education level implied by estimates in Table S3 in Appendix D. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). The two figures on the left show correlations between salaried abilities (white collar “wc”, blue collar “bc”) and each of the entrepreneurial abilities. Third figure shows the correlation between entrepreneurial abilities. Fourth figure shows the correlation between salaried abilities.

comes (Table S4 in Appendix D), which can slow down learning. As an assessment of how fast own- and cross-occupation learning about ability can happen, Figure 8 displays the percent of prior uncertainty about entrepreneurial ability that is eliminated after working for 5 years in each occupation. For individuals with more than college education, uncertainty about incorporated ability decreases by almost 90 percent after 5 years of incorporated experience and decreases by about 30 percent after 5 years of white collar experience. Surprisingly, for the college educated, paid employment is a slightly better source for learning about incorporated ability than incorporated entrepreneurship itself. The same result does not hold for unincorporated entrepreneurship.



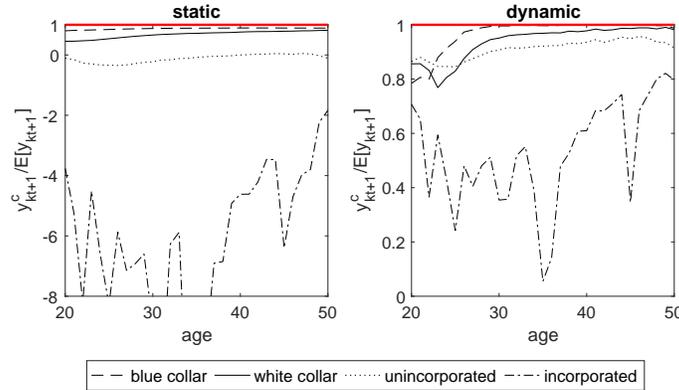
**FIGURE 8: Prior Variance Eliminated After 5 Years of Occupation-Specific Experience**

Notes: Percent of prior variance about ability in occupation  $k \in \{\text{unincorporated, incorporated}\}$  eliminated after accumulating five years of experience in each occupation  $k'$  and zero years in  $k'' \neq k'$ . Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). Computations use estimates in Tables S3 and S4 in Appendix D.

## 5.2 Risk Aversion, Costs and Benefits

This section discusses the estimates of risk aversion, entry costs and non-pecuniary benefits. In presenting results, the list of point estimates is deferred to Table S5 in Appendix D, and utility parameters are discussed using a certainty equivalent exercise and monetary equivalents for entry costs and non-pecuniary benefits.<sup>37</sup>

Starting with risk aversion, Figure 9 shows the ratio of certainty equivalent to expected income. Taking the estimates as given, the left panel assumes myopic individuals and the right panel considers forward-looking individuals. The certainty equivalent for entrepreneurship is much lower (in fact, negative) for myopic individuals than forward-looking individuals, which reflects high potential losses due to the high variance in entrepreneurial ability, especially for incorporated entrepreneurship. Participation in entrepreneurship, in the face of risk aversion, requires from entrepreneurs to value expected benefits associated to future human capital and future information. This dynamic consideration mitigates the effects of risk aversion.



**FIGURE 9:** Certainty Equivalent

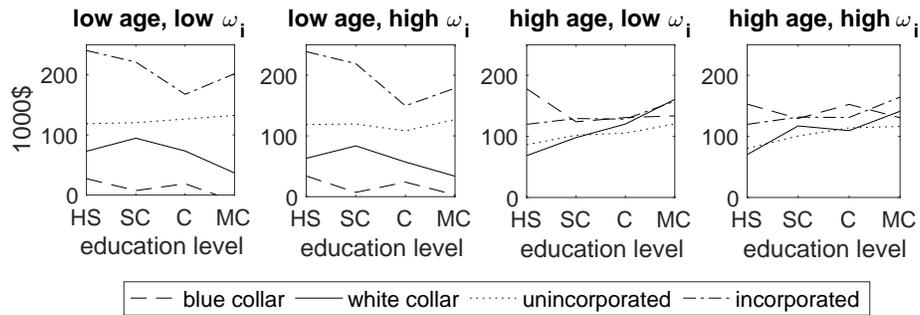
Notes: On the y-axis is the average of the ratio of the certainty equivalent  $y_{kt+1}^c$  (Appendix D) to the expected hourly income conditional on beliefs,  $E[y_{kt+1}|h_t, \mathbb{E}_t]$ , computed across individuals of a given age with positive expected income.

In addition to risk aversion, first-time entry costs constitute a barrier to entrepreneurship, in particular to young entrepreneurship. Figure 10 introduces monetary equivalents of entry cost estimates.<sup>38</sup> On the one hand, the monetary equivalent of entry cost into incorporated entrepreneurship

<sup>37</sup> Risk aversion and most entry cost and non-pecuniary benefit parameters are statistically significant.

<sup>38</sup> In the specification of the model, entry costs and non-pecuniary benefits can be treated as the indirect utility

is about \$50,000 higher for individuals age 20 than for individuals age 40. On the other, the entry cost into incorporated entrepreneurship is only about \$20,000 higher for individuals in the 10th percentile of permanent wealth versus individuals in the 90th percentile.<sup>39</sup> As expected, there is a negative relation between entrepreneurial entry costs and permanent wealth. The steep 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.<sup>40</sup>



**FIGURE 10: Monetary Equivalent of Entry Costs**

Notes: On the x-axis of each panel is the education level: 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 S5 and equation (S58) both in Appendix D. Ages are 20 and 40. Permanent wealth  $\omega_i$  points are the 10th and 90th percentiles.

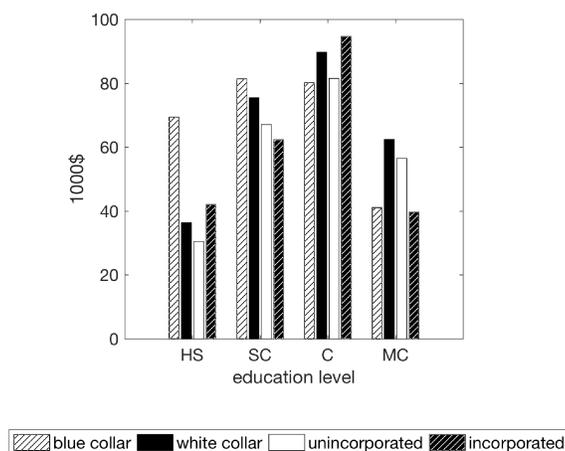
Non-pecuniary motivations can also affect the decision to become an entrepreneur (Hamilton, 2000; Hurst and Pugsley, 2015). Monetary equivalents of non-pecuniary benefits in Figure 11 suggest that entrepreneurial non-pecuniary benefits do not always dominate salaried non-pecuniary benefits once dynamic considerations are introduced.<sup>41</sup> Entrepreneurship is ranked below blue collar work by about \$20,000 per year for low educated individuals. However, for individuals with college or more, entrepreneurship becomes more attractive. In particular, the non-pecuniary benefits of incorporated entrepreneurship for the college educated are higher than those from any representation of cost terms in the budget constraint (Appendix D).

<sup>39</sup> A counterfactual exercise presented in Table S11 in Appendix D shows that eliminating all variation in permanent wealth by setting all individual's permanent wealth at the median has virtually no effects on entry and age at first entry into entrepreneurship.

<sup>40</sup> The dispersion in entry costs across-occupations in Figure 10 decreases with age as entrepreneurial entry costs decline and salaried entry costs increase. This may result from older individuals having a harder time starting careers in paid employment due to difficulties in obtaining or regaining skills at old ages.

<sup>41</sup> These results are in line with those in Dillon and Stanton (2017).

other occupation at any education level.



**FIGURE 11: 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). On the  $y$ -axis is the monetary equivalent of the non-pecuniary benefits not related to entry, obtained for a white, married man using the estimates in Table S5 in Appendix D and a similar derivation as in equation (S58) in Appendix D.

The treatment of the entrepreneurial choice in this paper captures many of its economic determinants. However, there are at least two caveats of this analysis. First, due to data limitations and model tractability, credit constraints are not modeled explicitly. Consistent with the credit constraints hypothesis, however, results show that 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.

## 6 The Determinants of Entrepreneurial Choice

Section 2 showed that most individuals do not attempt entrepreneurship and that those who do, enter later in their careers. This section revisits these stylized facts quantifying the relative importance of the various mechanisms in the model. The quantification relies on a comparison of the baseline model against counterfactual regimes that disable the mechanisms.<sup>42</sup> Results below include

<sup>42</sup> The comparison requires solving the model and simulating data under the baseline and each of the counterfactuals, keeping the initial state is fixed. Appendix D shows how the model is solved backwards using the representation in Proposition 2, and further explains the counterfactual regimes including extended results from the comparison among them in Table S11.

both types of entrepreneurs, but the discussion focuses on incorporated entrepreneurs. Information frictions and risk aversion play an important role in preventing participation in incorporated entrepreneurship, and the main forces explaining the gap in first entry ages are entry costs and lack of information. In addition, this section studies the effects of the mechanisms on long term outcomes focusing on the present value of income (PVI). Information frictions also have a large effect in the long-term: fully-informed incorporated entrepreneurs have a PVI that is about 50% higher than in the baseline. Moreover, removing the ability to use paid-employment outcomes to predict incorporated entrepreneurial ability decreases PVI for incorporated entrepreneurs by about 25%.

The counterfactual regimes used for decomposition are:

- ↔ *No learning-by-doing*: productivity does not increase with experience. Instead, occupational skill is constant and pays an average return.<sup>43</sup>
- ↔ *No learning about ability (full information)*: individuals know their ability vector  $\mathcal{M}_i$  but the initial level of uncertainty remains unchanged.<sup>44</sup>
- ↔ *No cross-occupation learning-by-doing*: productivity in one occupation is invariant to experience in another.
- ↔ *No cross-occupation learning about ability (no correlated learning)*: individuals believe that their success in one occupation is uninformative of their ability in another. Their initial prior variance is  $\Delta_s$  diagonalized.

---

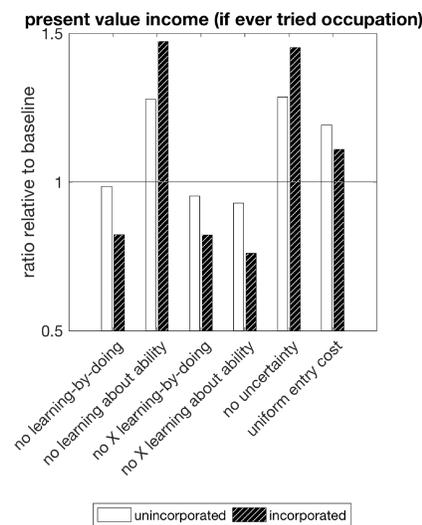
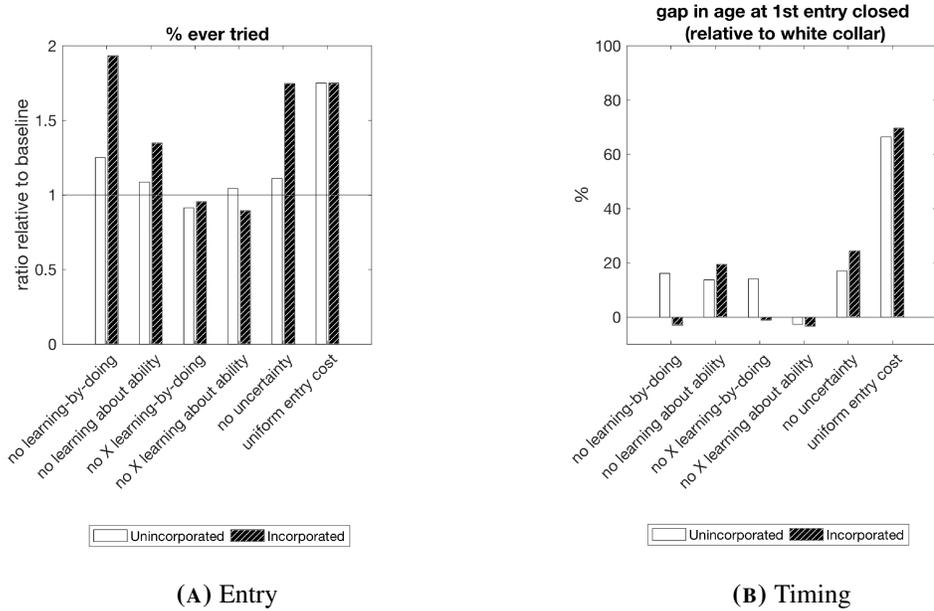
<sup>43</sup> Let  $R_k(x)$  be the marginal return to experience in occupation  $k$  for somebody who has worked  $x$  years in occupation  $k$  and zero years in any other occupation (Figure 6a). The fixed hourly return to observed skill for individuals in occupation  $k$  under this counterfactual regime is

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

<sup>44</sup> In terms of equation (2), this amounts to changing the value of the idiosyncratic income variance in occupation  $k$  from just  $\sigma_{\eta_k}$  to  $\sigma_{\eta_k} + \Delta_{s,\{k,k\}}$ , where  $\Delta_{s,\{k,k\}}$  is the  $k$ -th term in the diagonal of  $\Delta_s$ .

↪ *No uncertainty*: individuals know their ability vector  $\mathcal{M}_i$  and there is no idiosyncratic variation around their hourly income.

↪ *Uniform entry cost*: entry cost does not vary with age. Instead, individuals pay the cost faced by a 35-year old individual with their same education level.



**FIGURE 12: The Determinants of Entrepreneurial Choice**

Notes: Comparison of simulated data from the baseline model versus several counterfactual regimes. The labels “no X learning-by-doing” and “no X learning about ability” stand for no cross-occupation returns to experience and no correlated learning about ability, respectively.

Several forces prevent individuals from attempting entrepreneurship: having to learn how to be one, uncertainty in ability, risk aversion and entry costs. Figure 12a displays the ratio of the share of individuals who attempt entrepreneurship during their careers in each of the counterfactuals relative to the baseline. Disabling learning-by-doing has a strong effect on incorporated entry. If people did not have to learn-by-doing their way up through the incorporated productivity ladder (Figure 6a), they would be almost twice as likely to attempt incorporated entrepreneurship. Risk aversion and imperfect information also play important roles. Under full information about ability 35% more individuals attempt incorporated entrepreneurship. The effect of risk aversion is measured as the difference in ratios from providing full information versus eliminating all uncertainty, which amounts to about 40% for incorporated entrepreneurship.<sup>45</sup> Flattening the entry cost profile has a similar effect to eliminating all uncertainty: individuals would be 75% more likely to attempt incorporated entrepreneurship.<sup>46</sup> Cross-occupation learning has a positive impact on experimentation. Once cross-occupation learning is disabled, those who were set to gain the most from switching in the baseline (white collar workers) are no longer able to improve neither their entrepreneurial productivity nor their entrepreneurial beliefs using their paid-employment experience and outcomes.

Young entrepreneurs are harder to find because young individuals face higher entry costs and they have less information about their entrepreneurial ability. Table 1 in Section 2 showed the gap in first-entry ages between entrepreneurial occupations and salaried occupations. Figure 12b shows the percentage of the gap, relative to white collar, that is closed under each of the counterfactual regimes. Flattening the entry cost profile closes 70% of the gap and providing full information about ability closes 20%. Risk aversion has a smaller effect on the gap in first entry ages, as

---

<sup>45</sup> Recall that in the full information counterfactual individuals know their ability vector but the initial level of uncertainty remains unchanged. This is how the effect of risk aversion is separated from the effect of sorting on ability.

<sup>46</sup> Given the results in Table S5 in Appendix D, flattening the entry costs with respect to age amounts to young individuals facing a lower entry cost than in the baseline. For comparison, Table S11 in Appendix D includes a counterfactual in which permanent wealth is set at the 99th percentile for everyone. In this case, individuals are 33% more likely to attempt incorporated entrepreneurship.

illustrated by the additional 5% reduction of the gap on top of the decrease attained from full information. Interestingly, eliminating correlated learning widens the gap by about 3%. In the absence of correlated learning about ability individuals who start their careers as paid employees do not update entrepreneurial beliefs, as a consequence they postpone their first entry into entrepreneurship until their paid-employment beliefs have deteriorated enough.

Finally, as a measure of the long-term effects of the economic forces on the outcomes of entrepreneurs, Figure 12c shows the present value of the entrepreneurs' realized stream of income relative to the baseline. The effect of imperfect information is large: fully informed incorporated entrepreneurs have a present value of income that is about 50% higher. Flattening entry costs also increases the PVI of the incorporated, although only by about 10%, because successful young entrepreneurs that decide to stay will enjoy the returns of their high ability for longer. Shutting down all learning-by-doing or just cross-occupation learning-by-doing both reduce the PVI by about 20%. It follows that the main reason behind this decline is the lack of transferability of experience. Notably, the long-term effect of using paid-employment outcomes to predict incorporated entrepreneurial success is not negligible. Disabling correlated learning decreases the incorporated PVI by about 25%.

## **7 Fostering Young Entrepreneurship**

Section 6 showed that the main barriers to young entrepreneurship are entry costs and information frictions. In this section, counterfactual policies focusing on incorporated entrepreneurship target these barriers. A blanket subsidy increases entrepreneurship but has a small long-term effect as measured by the present value of income. Entrepreneurship education, characterized as a source of information, can have a sizable effect on participation and the present value of income. I complement these results using the structural framework to show that one motivation for promoting young entrepreneurship lies in the information value of early signals, given the high variance in entrepreneurial ability. I show that the gains from entrepreneurship education programs decrease

rapidly with age.

## 7.1 Subsidies

A subsidy for young incorporated entrepreneurs can foster participation by targeting the high entry cost they face. In this counterfactual individuals receive either \$25,000 or \$50,000 if they decide to start their careers as incorporated entrepreneurs immediately after finishing their education. The *Subsidy* columns of Table 6 summarize the effects of the intervention in terms of the pool of young incorporated entrepreneurs, the overall pool of incorporated entrepreneurs, and the PVI of all individuals in the economy—the last two capture 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%.

The subsidy may be justified regardless of this adverse effect on the quality of young incorporated entrepreneurs. For instance, the subsidy may attract individuals of exceptional ability 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. In the long term, the \$50,000 subsidy increases by 50% the number of people who attempt incorporated entrepreneurship in their careers. Besides, the average net present value of income of all individuals in the economy (including those who do not experiment with entrepreneurship) increases by about 3%. This happens because the distribution of entrepreneurial ability is wider than the distribution of paid-employment ability. Hence, the entrepreneurial ability of the marginal entrepreneur can generate more hourly income than his correspondent ability in paid employment.

Focusing on the average entrepreneurial ability of new entrants, results here seem consistent with those in Hamilton et al. (2018) and underscore the arguments 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 harmless. In fact, although small, there may be gains in terms of PVI from young successful entrepreneurs attracted early on by a subsidy.

**TABLE 6: Policies Fostering Young Incorporated Entrepreneurship**

	<i>Baseline</i>	<i>Subsidy</i>		<i>Entrepreneurship Education</i>			
		\$1000s		noise variance scale ( <i>s</i> )			
		25	50	10	5	2	1
<b>Young Entrepreneurs</b>							
Tried in first 5 years	0.02	0.03	0.05	0.12	0.11	0.10	0.08
Mean belief (\$ per hour) at 1st entry	3.7	2.5	1.2	82.4	75.1	61.3	48.4
Mean ability (\$ per hour) at 1st entry	5.0	3.6	2.1	1.3	2.1	5.5	10.7
Bias (belief-ability)	-1.3	-1.1	-0.9	81.1	73.0	55.8	37.7
95th pctile ability (\$ per hour) at 1st entry	46.5	44.0	37.8	46.9	47.5	49.9	59.1
<b>Overall</b>							
Tried	0.15	0.17	0.22	0.26	0.26	0.25	0.24
Participation rate at age 40	0.04	0.05	0.06	0.12	0.12	0.12	0.11
PVI (\$1000s)	757	770	770	944	941	961	983
Mean belief (\$ per hour) at 1st entry	6.4	6.2	5.2	45.9	41.8	33.8	26.5
Mean ability (\$ per hour) at 1st entry	5.5	5.4	4.6	2.0	2.3	4.0	6.6
Bias (belief-ability)	0.9	0.8	0.6	43.9	39.5	29.8	19.9
<b>All individuals</b>							
PVI (\$1000s)	508	513	526	573	575	578	581

Notes: *Rows*: summary statistics provided separately for young incorporated entrepreneurs, all incorporated entrepreneurs, and all individuals. Young entrepreneurs are those who try incorporated entrepreneurship for the first time within their first five years in the labor market. PVI stands for the present value of income and is computed net of the subsidy for the subsidy columns. *Baseline* column: the baseline model where there are no subsidies or education. *Subsidy* columns: subsidy of \$25,000 or \$50,000 given only to individuals who attempt incorporated entrepreneurship immediately after finishing their education. *Entrepreneurship Education* columns: individual-specific signal about incorporated ability given to everybody immediately after finishing their education. Interventions are characterized by the noise variance of their signals,  $\sigma_v$ , expressed as a scaled version of the noise variance of trying incorporated entrepreneurship in reality:  $\sigma_v^2 = s \cdot \sigma_{\eta_4}^2$ .

## 7.2 Entrepreneurship Education

Many policies that attempt to foster young entrepreneurship focus on entrepreneurship education.<sup>47</sup> To the extent that these policies help reveal entrepreneurial potential, the emphasis on entrepreneurship education is consistent with the results in Section 6 showing that information frictions represent an important barrier to young entrepreneurs. Previous literature suggests that entrepreneurship education programs can shift individuals' elicited beliefs and intentions (Souitaris et al., 2007; von Graevenitz et al., 2010; Oosterbeek et al., 2010), but the value of these policies

<sup>47</sup> 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.

critically depends on the quality of the information they provide. I extend the literature by providing a mapping from entrepreneurship education of a given quality, through shifts in beliefs, into career choices and long-term outcomes. In other words, I provide a dynamic framework to assess the value of any given policy that provides information. Using findings in the literature, I undertake a calibrating exercise that pinpoints the position of an entrepreneurship program in the information quality spectrum, providing an opportunity to assess the long-term effects of similar entrepreneurship education programs.

Entrepreneurship education is characterized here as a source of information. Before entering the job market, 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 counterfactual policy effectively induces initial heterogeneity in entrepreneurial beliefs that will depend on ability as well as on luck. Notably, because abilities are correlated, this policy induces heterogeneity in beliefs across all occupations. Formally, the entrepreneurship education program yields every individual a signal  $\zeta_i^p$  about his incorporated ability ( $\mu_{4,i} \in \mathcal{M}_i$ ) given by

$$\zeta_i^p = \mu_{4,i} + v_i \quad (12)$$

where  $v_i$  are iid  $N(0, \sigma_v^2)$ . Individuals use the information contained in  $\zeta_i^p$  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 for one period. In other words, the noise variance from this intervention is bounded below by the estimated idiosyncratic variance  $\sigma_{\eta_4}^2$  (Table S4 in Appendix D). Therefore, the noise variance from the entrepreneurship education program can be written as

$$\sigma_v^2 = s \cdot \sigma_{\eta_4}^2, \quad \text{with } s \geq 1 \quad (13)$$

The *Entrepreneurship Education* columns of Table 6 present the effects of policies that differ in their information quality ( $s \in \{1, 2, 5, 10\}$ ). The lower the quality (higher  $s$ ), the higher the percent-

age of young entrepreneurs and the bias in their beliefs. Table 6 also shows a decline in average ability of young entrepreneurs for information quality below 50% ( $s = 2$ ). This reflects the number 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 entrepreneurship education that provides 10% information quality ( $s = 10$ ). Although these results relate to previous literature suggesting that overconfidence influences entrepreneurial entry (Camerer and Lovallo, 1999), overestimation of ability in my framework is not a different psychological trait of entrepreneurs or the result of differential analysis of the information received (March and Shapira, 1987). Instead, overestimation at first entry emerges endogenously from uninformed rational individuals who are fortunate to receive large positive signals.

Entrepreneurship education can also provide long-term gains even when the quality of information is as low as 10%. Table 6 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, entrepreneurship education could benefit all individuals, not only those who eventually become entrepreneurs. According to the last row in Table 6, the average individual's gain from the program relative to the baseline vary from \$65,000 when  $s = 10$  to \$73,000 when  $s = 1$ .<sup>48</sup>

Putting these results in perspective, I use findings in the literature to calibrate the information quality of one particular entrepreneurship program. Concretely, von Graevenitz et al. (2010) show that the standard deviation of aggregate beliefs for college students increases as a consequence of the entrepreneurship education program they study. I use my model to obtain a mapping from information quality of a program—characterized by  $s$ —into changes in the standard deviation of aggregate beliefs, and calibrate the value of  $s$  to match the change in aggregate beliefs found

---

<sup>48</sup> To reflect the differences in entrepreneurial potential at every education level, these quantities can also be computed by education level. The difference in expected PVI from entrepreneurship education of 10% of quality goes from \$700 for high schoolers to about \$200,000 for individuals with more than college education.

in their paper (Appendix D). This exercise indicates that the information from the program is around half the quality of the information in the market ( $s = 2.1$ ). Using Table 6 for reference, a program of similar information quality ( $s = 2$ ) would increase young incorporated entrepreneurship from 2% to approximately 10%, raise the percent of people who experiment with incorporated entrepreneurship in their careers from 15% to about 25%, and add \$70,000 to the PVI of the average individual.

### 7.3 Why Young Entrepreneurs?

Although well-known international institutions have discussed the need to foster young entrepreneurship (OECD, 2013), it is unclear what are the returns of this approach relative to fostering entrepreneurship at other stages in a person's career.<sup>49</sup> I provide an explanation for this approach using the entrepreneurship education policy from the previous section. Setting the quality of information to the calibrated value ( $s = 2.1$ ), I compare the intervention's results when introduced (unexpectedly) at three different stages in an individual's career (Table 7): at the beginning, at age 30, and at age 40. Even though the share of individuals who try incorporated entrepreneurship remains almost unchanged with the timing of the intervention, the gains from the intervention in terms of PVI decrease rapidly with its timing. When introduced at age 40, the intervention only generates 17% of the additional \$74,000 in PVI generated when the intervention is at the beginning of people's careers. Hence, the model offers a simple economic reason to foster young entrepreneurship: the value of early information is much larger.

## 8 Conclusion

On the basis of its potential economic benefits, entrepreneurship is the target of policy interventions at different stages in the life cycle. In the data most people do not attempt entrepreneurship during their careers, and those who attempt it do it after going through several years in paid employment.

---

<sup>49</sup> Young entrepreneurship is often motivated as an avenue to avoid unemployment.

**TABLE 7: Policies Fostering Young Incorporated Entrepreneurship**

	<i>Entrepreneurship Education</i>			
	<i>Baseline</i>	<i>Beginning</i>	<i>Age 30</i>	<i>Age 40</i>
<b>Overall</b>				
Tried	0.15	0.25	0.25	0.24
PVI (\$1000s)	508	582	544	521

Notes: *Rows*: summary statistics provided for all individuals. PVI stands for the present value of income. *Baseline* column: the baseline model where there is no intervention. *Entrepreneurship Education* columns: individual-specific signal about incorporated ability given to everybody. *Beginning* column: intervention immediately after finishing their education. *Age AA* column: intervention at age *AA*. Intervention characterized by the noise variance of its signals,  $\sigma_v$ , expressed as a scaled version of the noise variance of trying incorporated entrepreneurship in reality:  $\sigma_v^2 = s \cdot \sigma_{\eta_4}^2$ , where  $s = 2.1$ .

I study the mechanisms behind these stylized facts thereby exploring the channels through which entrepreneurship policies act. Among the mechanisms I study are information frictions, cross-occupation learning, entry costs and risk aversion—which I show plays an important role and has been overlooked in dynamic, empirical models of entrepreneurship. In the paper I not only provide the determinants of the stylized facts but also a framework for evaluating entrepreneurship policies. In particular, I show that policies that provide information could generate larger effects on the present value of life-time income, and I provide an explanation for why it makes sense for policy makers to focus on young entrepreneurship: the value of early signals is larger.

Many policies fostering entrepreneurship not only seek to attract new participants but also to increase employment through the jobs new entrepreneurs can create. Since it is not clear whether these interventions affect the decision to hire employees, more work is needed to evaluate these policies taking into account not only the number and quality of new entrepreneurs, but also their propensity to generate jobs. The framework introduced here is a 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 by new entrepreneurs.

## References

- AGUIRREGABIRIA, V. AND P. MIRA, “Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models,” *Econometrica* 70 (2002), 1519–1543.
- ALTUČ, S. AND P. LABADIE, *Dynamic Choice and Asset Markets* (Academic Press, 1994).
- AMEMIYA, T., *Advanced Econometrics* (Harvard University Press, 1985).
- ANTONOVICS, K. AND L. GOLAN, “Experimentation and Job Choice,” *Journal of Labor Economics* 30 (2012), 333–366.
- ARCIDIACONO, P. AND R. A. MILLER, “Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity,” *Econometrica* 79 (2011), 1823–1867.
- , “Identifying Dynamic Discrete Choice Models off Short Panels,” Working Paper, 2015.
- BUERA, F. J., “A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence,” *Annals of finance* 5 (2009), 443–464.
- CAMERER, C. AND D. LOVALLO, “Overconfidence and Excess Entry: An Experimental Approach,” *The American Economic Review* 89 (1999), 306–318.
- DILLON, E. W. AND C. T. STANTON, “Self-Employment Dynamics and the Returns to Entrepreneurship,” Nber working Paper 23168, 2017.
- EVANS, D. S. AND B. JOVANOVIĆ, “An Estimated Model of Entrepreneurial Choice under Liquidity Constraints,” *The Journal of Political Economy* (1989), 808–827.
- GAYLE, G.-L., L. GOLAN AND R. A. MILLER, “Promotion Turnover and the Compensation in the Executive Labor Market,” *Econometrica* 83 (2015), 2293–2369.
- GIBBONS, R., L. F. KATZ, T. LEMIEUX AND D. PARENT, “Comparative Advantage, Learning, and Sectoral Wage Determination,” *Journal of Labor Economics* 23 (2005), 681–724.

- GREEN, E. J., “Lending and the Smoothing of Uninsurable Income,” in E. Prescott and N. Wallace, eds., *Contractual Arrangements for Intertemporal Trade* chapter 1 (University of Minnesota Press, 1987), 3–25.
- HALL, R. E. AND S. E. WOODWARD, “The Burden of the Nondiversifiable Risk of Entrepreneurship,” *The American Economic Review* 100 (2010), 1163–1194.
- HAMILTON, B. H., “Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment,” *The Journal of Political Economy* 108 (2000), 604–631.
- HAMILTON, B. H., N. PANDE AND N. W. PAPAGEORGE, “The Right Stuff? Personality and Entrepreneurship,” Accepted at *Quantitative Economics*, 2018.
- HECKMAN, J. J. AND B. E. HONORÉ, “The Empirical Content of the Roy Model,” *Econometrica* 58 (1990), 1121–1149.
- HOTZ, V. J. AND R. A. MILLER, “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *The Review of Economic Studies* 60 (1993), 497–529.
- HUMPHRIES, J. E., “The Causes and Consequences of Self-Employment over the Life Cycle,” Mimeo, Yale University, 2018.
- HURST, E. AND A. LUSARDI, “Liquidity constraints, household wealth, and entrepreneurship,” *The Journal of Political Economy* 112 (2004), 319–347.
- HURST, E. AND B. W. PUGSLEY, “Wealth, Tastes, and Entrepreneurial Choice,” Working Paper, NBER, 2015.
- IYIGUN, M. F. AND A. L. OWEN, “Risk, Entrepreneurship, and Human-Capital Accumulation,” *The American Economic Review* 88 (1998), 454–457.
- JAMES, J., “Ability Matching and Occupational Choice,” SSRN Scholarly Paper ID 1937879, 2011.

- JOVANOVIC, B., “Job Matching and the Theory of Turnover,” *The Journal of Political Economy* 87 (1979), 972–990.
- JOVANOVIC, B. AND Y. NYARKO, “Learning by Doing and the Choice of Technology,” *Econometrica* 64 (1996), 1299–1310.
- KEANE, M. P. AND K. I. WOLPIN, “The Career Decisions of Young Men,” *The Journal of Political Economy* 105 (1997), 473–522.
- LAFONTAINE, F. AND K. SHAW, “Serial Entrepreneurship: Learning by Doing?,” *Journal of Labor Economics* 34 (2016), S217–S254.
- LAZEAR, E. P., “Entrepreneurship,” *Journal of Labor Economics* 23 (2005), 649–680.
- LEVINE, R. AND Y. RUBINSTEIN, “Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?,” *The Quarterly Journal of Economics* 132 (2017), 963–1018.
- LINDQUIST, M. J., J. SOL AND M. VAN PRAAG, “Why Do Entrepreneurial Parents Have Entrepreneurial Children?,” *Journal of Labor Economics* 33 (2015), 269–296.
- MACURDY, T. E., “An Empirical Model of Labor Supply in a Life-Cycle Setting,” *The Journal of Political Economy* 89 (1981), 1059–1085.
- MAGNAC, T. AND D. THESMAR, “Identifying Dynamic Discrete Decision Processes,” *Econometrica* 70 (2002), 801–816.
- MANSO, G., “Experimentation and the Returns to Entrepreneurship,” *The Review of Financial Studies* 29 (2016), 2319–2340.
- MARCH, J. G. AND Z. SHAPIRA, “Managerial Perspectives on Risk and Risk Taking,” *Management Science* 33 (1987), 418–438.
- MARGIOTTA, M. M. AND R. A. MILLER, “Managerial Compensation and the Cost of Moral Hazard,” *International Economic Review* 41 (2000), 418–438.

- MILLER, R. A., “Job Matching and Occupational Choice,” *The Journal of Political Economy* 92 (1984), 1086–1120.
- OECD, “Supporting Youth in Entrepreneurship,” Seminar Summary, OECD, Paris, 2013.
- OOSTERBEEK, H., M. VAN PRAAG AND A. IJSSELSTEIN, “The Impact of Entrepreneurship Education on Entrepreneurship Skills and Motivation,” *European Economic Review* 54 (2010), 442–454.
- SCHUMPETER, J. A., *The Theory of Economic Development* (Cambridge, MA: Harvard University Press, 1911).
- SOUITARIS, V., S. ZERBINATI AND A. AL-LAHAM, “Do Entrepreneurship Programmes Raise Entrepreneurial Intention of Science and Engineering Students? The Effect of Learning, Inspiration and Resources,” *Journal of Business Venturing* 22 (2007), 566–591.
- VON GRAEVENITZ, G., D. HARHOFF AND R. WEBER, “The Effects of Entrepreneurship Education,” *Journal of Economic Behavior & Organization* 76 (2010), 90–112.

For Online Publication Supplement to  
“Entrepreneurship over the Life Cycle:  
Where are the Young Entrepreneurs?”

## 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, wealth, working hours, completed education, age, race, and marital status.

*PotentialExperience*. In order to avoid integration over long sequences of unobserved income signals, which becomes too burdensome, I only keep individuals who are observed from the beginning of their labor market careers. For this purpose potential experience is defined as

$$PotentialExperience = Age - CompletedEducation - 6 \quad (S1)$$

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

$$PotentialExperience = \begin{cases} 0 & \text{if } \min PotentialExperience \leq 0 \\ k & \text{if } \min PotentialExperience = k \in \{1, 2, 3\} \end{cases} \quad (S2)$$

*Self-employment.* At any period, conditional on having declared to be working, working for money, or only temporarily laid off, individuals answer a version of the following question: “*On your main job, are you self-employed, are you employed by someone else, or what?*.” The answer options are “Someone else,” “Both someone else and self,” “Self-employed only,” and “Don’t Know.” Entrepreneurs are 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 three-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. After dropping observations corresponding to members of the armed forces, farm related occupations, and private household workers, 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

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.

*Unemployed status.* Individuals are classified as unemployed if they reported to be not working or working for less than 2.5% the total number of available hours in a year.

*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.

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

- ↔ 1. They answer “yes” to a version of the following question: “*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 affirm that the business mentioned was not uniquely a corporation. In other words, they 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 then they answer the following 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 for unincorporated self-employed individuals, who are the only ones who report it in this way.

*Income.* In summary, for salaried workers and incorporated self-employed individuals income equals reported labor income. For self-employed unincorporated individuals income equals reported labor income plus reported business income. 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. Additionally, 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.

The algorithm seeks to provide 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 in IQ1 (individual reported "Both," "Other," or "Do not know"), the value from question IQ2 is assigned insofar as it corresponds to "Incorporated" or "Unincorporated;" **(iii)** if data is still missing or ambiguous, past and future answers ( $t - 5, \dots, t - 1, t + 1, \dots, t + 5$ ) 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 between steps **(ii)** and **(iv)**: 406 from step **(ii)**, 120 from step **(iii)** and 25 from step **(iv)**.

*Working Hours.* Individuals report hours worked during year YYYY. Missing data were not assigned.

*Hourly income.* Computed 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).

*Experience variables.* 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, I use the observations available on wealth to construct a measure of permanent wealth. In order to obtain the individual measure I run the following fixed effects regression:

$$Wealth_{it} = \gamma_0 + \gamma_1 age_{it} + \gamma_2 age_{it}^2 + u_i + \varepsilon_{it} \quad (S3)$$

Using results in Table S1, the individual permanent wealth measure is obtained as

$$\omega_i = \hat{\gamma}_0 + \hat{u}_i \quad (S4)$$

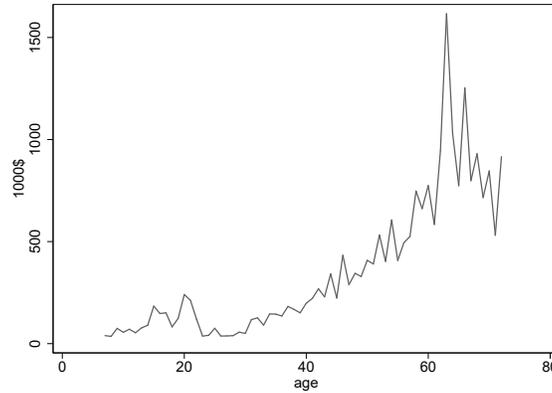
In estimation only individuals with at least three wealth data points are included. Figure S1 shows the age profile of wealth accumulation from pooling all available data.

**APPENDIX TABLE S1:** Parameters of the Wealth Profile Equation

	coeff	se
$\gamma_0$	417.78	(85.16)
$\gamma_1$	-23.95	(4.40)
$\gamma_2$	0.47	(0.06)

Notes: wealth in thousands of dollars of 2000.

*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 the data, only about 6% of individual-year observations for working individuals are part-time observations (less than 20 hours per week). Part-time



**APPENDIX FIGURE S1: Average Wealth**

Notes: Measured in thousands of dollars of 2000.

individual-year observations are not dropped because they would create gaps in the careers of 35% of the individuals.

*Data Gaps.* the histories of individuals with data gaps wider than 2 years are dropped. For those with data in  $t$  and  $t + 2$  but not in  $t + 1$ , time is 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;<sup>50</sup> 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 observation from the beginning of their careers: 37,759. (This criterion, and the one for wealth data below, create a disparity between the proportion of individuals with at least college education in the final sample (0.42) and the proportion for the U.S.

<sup>50</sup> Relevant data on income and occupation is only collected for household heads.

adult population during the period (around 0.22) shown in Ryan and Siebens (2012). The disparity arises because the lower an individual's education level is, the less likely the PSID is to observe him from the beginning of his career.) 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. This yields  $b_t$  defined as

$$\begin{aligned}
 b_t &\equiv \sum_{s=1}^{\infty} \prod_{i=1}^s (1 + r_{it})^{-1} \\
 &= \sum_{s=1}^{30} \prod_{i=1}^s (1 + r_{it})^{-1} + \prod_{i=1}^{30} (1 + r_{it})^{-1} \sum_{s=31}^{\infty} (1 + r_{30,t})^{s-30} \\
 &= \sum_{s=1}^{30} \prod_{i=1}^s (1 + r_{it})^{-1} + \frac{1}{r_{30,t}} \prod_{i=1}^{30} (1 + r_{it})^{-1}
 \end{aligned} \tag{S5}$$

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  $i \in \{1, \dots, 30\}$ .

*Step 1:* use the annual compounding interest rate  $\tilde{r}_{st}$  (from the interpolated yield curve) to obtain  $b_t$  as

$$b_t = \sum_{s=1}^{30} \left( \frac{1}{1 + \tilde{r}_{st}} \right)^s + \frac{1}{r_{30,t}} \left( \frac{1}{1 + \tilde{r}_{30,t}} \right)^{30} \quad (\text{S6})$$

*Step 2:* given that  $r_{it}$  and  $\tilde{r}_{it}$  are nominal interest rates, adjust  $b_t$  by the deflator based on base year 2000 to reflect inflation

$$\tilde{b}_t = \frac{b_t}{\text{deflator}_{2000}} \quad (\text{S7})$$

The series of  $\tilde{b}_t$  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 in-sample bond prices  $\tilde{b}_t$  for  $t = 1954, \dots, 2015$ , a regression is run for  $\tilde{b}_{t+1}$  on  $\tilde{b}_t$  to obtain out-of-sample prices  $\hat{b}_t$  for  $t = 2016, \dots, 2033$ .

## B Model Appendix

### B.1 Proof of Proposition 1

*Proof:* The proof works by backwards induction. 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

$$V_{T+1}(h_{T+1}, \mathbb{B}_{T+1}, \xi_{T+1}, a_{\tau(T+1)}, b_{\tau(T+1)}) = -\lambda_{\tau(T+1)} b_{\tau(T+1)} \exp\left(\frac{-(\rho \xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right) \quad (\text{S8})$$

His occupational ability and his experience become irrelevant once he retires. Since 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  $b_{\tau(T+1)}$ .<sup>51</sup> Now consider 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

$$\begin{aligned} - \alpha_{Tk}(h_T) \beta^T \exp\{-\rho c_T - \varepsilon_{Tk}\} - E_T \left[ \lambda_{\tau(T+1)} b_{\tau(T+1)} v_{kT+1} \exp\left(\frac{-(\rho \xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right) \middle| \mathbb{B}_T, h_T \right] \\ \text{s.t. } E_T[\lambda_{\tau(T+1)} \xi_{T+1} | d_{Tk}, h_T, \mathbb{B}_T] + \lambda_{\tau(T)} c_T = \lambda_{\tau(T)} \xi_T \end{aligned} \quad (\text{S9})$$

His budget constraints in (S9) 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 (S8). Following a similar procedure as in Margiotta and Miller (2000, p. 680) the conditional value function of choosing alternative  $k$  is obtained as

$$\begin{aligned} V_{kT}(h_T, \mathbb{B}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}, \varepsilon_{kT}) = \\ -\lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} e^{-\varepsilon_{kT}/b_{\tau(T)}} E_T[v_{kT+1} | \mathbb{B}_T]^{1-1/b_{\tau(T)}} \exp\left(\frac{-(\rho \xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) \end{aligned} \quad (\text{S10})$$

Integrating over  $\varepsilon_T$  and averaging over the 5 choices using the conditional choice probabilities yields

---

<sup>51</sup> More complex models could have ability and accumulated human capital generating an income stream after retirement. I abstract from such considerations acknowledging that retirement could play an important role if occupational paths that include entrepreneurship generate very different retirement income flows.

$$\begin{aligned}
V_T(h_T, \mathbb{B}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}) &= - \sum_{k=0}^4 p_{kT}(h_T, \mathbb{B}_T) \lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} E_\varepsilon[e^{-\varepsilon_{kT}/b_{\tau(T)}}] E_T[v_{kT+1} | \mathbb{B}_T]^{1-1/b_{\tau(T)}} \\
&\quad \times \exp\left(\frac{-(\rho \xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) \\
&= - \lambda_{\tau(T)} b_{\tau(T)} \exp\left(\frac{-(\rho \xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) A_T(h_T, \mathbb{B}_T)
\end{aligned} \tag{S11}$$

where

$$E_\varepsilon[e^{-\varepsilon_{kT}/b_{\tau(T)}}] \equiv E_\varepsilon[e^{-\varepsilon_{kT}/b_{\tau(T)}} | d_{kT} = 1] \tag{S12}$$

$A_T(h_T, \mathbb{B}_T)$  is defined as in equation (8) and  $A_{T+1}(h_{T+1}, \mathbb{B}_{T+1}) \equiv 1$ .

To finish the proof suppose that equations (7) and (8) hold for  $t + 1$ . Then, at age  $t$  an individual who has chosen alternative  $k$  selects consumption and savings to maximize

$$\begin{aligned}
& -\alpha_{kt}(h_t) \beta^t \exp\{-\rho c_t - \varepsilon_{kt}\} - E_t \left[ \lambda_{\tau(t+1)} b_{\tau(t+1)} A_{t+1}(h_{t+1}, \mathbb{B}_{t+1}) v_{kt+1} \exp\left(\frac{-(\rho \xi_{t+1} + a_{\tau(t+1)})}{b_{\tau(t+1)}}\right) \middle| \mathbb{B}_t, h_t, d_{kt} = 1 \right] \\
& s.t. \quad E_t[\lambda_{\tau(t+1)} \xi_{t+1} | d_{kt}, h_t, \mathbb{B}_t] + \lambda_{\tau(t)} c_t = \lambda_{\tau(t)} \xi_t
\end{aligned} \tag{S13}$$

Which yields an equation similar to equation (S10):

$$\begin{aligned}
V_{kt}(h_t, \mathbb{B}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}, \varepsilon_{kt}) &= \\
& - \lambda_{\tau(t)} b_{\tau(t)} \alpha_{kt}(h_t)^{1/b_{\tau(t)}} e^{-\varepsilon_{kt}/b_{\tau(t)}} E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1}) v_{kt+1} | \mathbb{B}_t, h_t]^{1-1/b_{\tau(t)}} \times \exp\left(\frac{-(\rho \xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right)
\end{aligned} \tag{S14}$$

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

## B.2 Proof of Proposition 2

*Proof:* Assuming that the taste shocks are distributed Extreme Value Type-I renders the expression in equation (9) as a standard logit. Hence, the odds ratio can be written as

$$\frac{p_{0t}(h_t, \mathbb{B}_t)}{p_{kt}(h_t, \mathbb{B}_t)} = \alpha_{kt}(h_t) E_t \left[ \frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})}{A_{t+1}(h_t, \mathbb{B}_t)} v_{kt+1} \Big| \mathbb{B}_t, h_t \right]^{b_{\tau(t)} - 1} \quad (\text{S15})$$

Equation (S15) 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 nor his beliefs change if he decides not to work. Use equation (S15) to write

$$E_t [A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1}) v_{kt+1} | \mathbb{B}_t, h_t]^{1-1/b_{\tau(t)}} = \alpha_{kt}(h_t)^{-1/b_{\tau(t)}} A_{t+1}(h_t, \mathbb{B}_t)^{1-1/b_{\tau(t)}} \left( \frac{p_{kt}(h_t, \mathbb{B}_t)}{p_{0t}(h_t, \mathbb{B}_t)} \right)^{-1/b_{\tau(t)}} \quad (\text{S16})$$

From the online appendix of Gayle et al. (2015, p. 3):

$$E_\varepsilon [e^{-\varepsilon_{kt}^*/b_{\tau(t)}}] = p_{kt}(h_t, \mathbb{B}_t)^{1/b_{\tau(t)}} \Gamma \left( \frac{b_{\tau(t)} + 1}{b_{\tau(t)}} \right) \quad (\text{S17})$$

where  $\Gamma(\cdot)$  denotes the complete gamma function. Substitute equations (S16) and (S17) in equation (8) to obtain

$$A_t(h_t, \mathbb{B}_t) = p_{0t}(h_t, \mathbb{B}_t)^{1/b_{\tau(t)}} \Gamma \left( \frac{b_{\tau(t)} + 1}{b_{\tau(t)}} \right) A_{t+1}(h_t, \mathbb{B}_t)^{1-1/b_{\tau(t)}} \quad (\text{S18})$$

Using equation (S18) we can write the ratio of indexes as

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})}{A_{t+1}(h_t, \mathbb{B}_t)} = \frac{p_{0t+1}(h_{kt}^{(1)}, \mathbb{B}_{kt}^{(1)})^{1/b_{\tau(t)+1}} A_{t+2}(h_{kt}^{(1)}, \mathbb{B}_{kt}^{(1)})^{1-1/b_{\tau(t)+1}}}{p_{0t+1}(h_{0t}^{(1)}, \mathbb{B}_{0t}^{(1)})^{1/b_{\tau(t)+1}} A_{t+2}(h_{0t}^{(1)}, \mathbb{B}_{0t}^{(1)})^{1-1/b_{\tau(t)+1}}} \quad (\text{S19})$$

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

$$\phi_t(s) = \frac{1}{b_{\tau(t)+s}} \prod_{r=1}^{s-1} (1 - 1/b_{\tau(t)+r}) \quad (\text{S20})$$

Iterative substitution of equation (S18) in (S19) up to retirement age yields

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})}{A_{t+1}(h_t, \mathbb{B}_t)} = \prod_{s=1}^{T-t} \left( \frac{p_{0t+s}(h_{kt}^{(s)}, \mathbb{B}_{kt}^{(s)})}{p_{0t+s}(h_{0t}^{(s)}, \mathbb{B}_{0t}^{(s)})} \right)^{\phi_t^{(s)}} \quad (\text{S21})$$

Plugging equation (S21) into equation (S15) and applying logarithms finishes the proof. *Q.E.D.*

### B.3 The Updating Rules

Updating rules for similar problems have been previously obtained in the literature (James, 2011). Define the  $K$ -dimensional vector  $\zeta_t$  with characteristic component  $\zeta_{\{k\}}$  and the  $K \times K$  diagonal matrix  $\Sigma_t$  with characteristic component  $\Sigma_{\{k,k\}it}$  as follows

$$\zeta_{\{k\}it} = \begin{cases} \zeta_{kt} & \text{if } d_{kt-1} = 1 \\ 0 & \text{otherwise} \end{cases} \quad \Sigma_{\{k,k\}it} = \begin{cases} 1/\sigma_{\eta_k}^2 & \text{if } d_{kt-1} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{S22})$$

After receiving his income signal from last period's work, his transition in beliefs described in equation (3) can be summarized by the updated mean  $\mathbb{E}_t$  and variance  $\mathbb{V}_t$  as follows:

$$\mathbb{E}_t = [\mathbb{V}_{t-1}^{-1} + \Sigma_t]^{-1} [\mathbb{V}_{t-1}^{-1} \mathbb{E}_{t-1} + \Sigma_t \zeta_t] \quad (\text{S23})$$

$$\mathbb{V}_t = [\mathbb{V}_{t-1}^{-1} + \Sigma_t]^{-1} \quad (\text{S24})$$

These updating rules reflect how beliefs change as a function of experience and information. Equations (S22) and (S23) imply that the effect of a very noisy signal on the prior mean is minor.<sup>52</sup> Notably, equation (S23) 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  $\mathbb{V}_t$ . The larger these covariances are, the larger the adjustment will be.<sup>53</sup> Equation (S24)

<sup>52</sup> A noisy signal is characterized by high idiosyncratic variance  $\sigma_{\eta_k}^2$ .

<sup>53</sup> The marginal effect of a signal from occupation  $k$  on the next period's prior mean of occupation  $k'$  equals  $(1/\sigma_{\eta_k}^2) \mathbb{V}_{\{k,k'\}t}$ .

implies that the prior variance at  $t$  is a deterministic map of the vector of accumulated experience  $x_t$  and the covariance matrix of the ability distribution. Hence, conditional on  $x_t$ , 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 the individual's state, equation (S24) implies that  $\mathbb{V}_t$  is redundant information.

## C Estimation Appendix

### C.1 The Likelihood

Rather than selecting on ability, individuals in the model select on beliefs. 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  $\Lambda$  be the collection of parameters of the utility function, let  $\Theta$  be the collection of income parameters, including the variance parameters of the productivity shocks,  $\sigma_{\eta_k}$ , and let  $\Delta_s$  be the covariance matrix of the population ability distribution conditional on education level  $s$ . Hence, the likelihood of the data—hourly income and choices—for a person with education level  $s$  can be written as:

$$\begin{aligned} \mathcal{L}_i &= \prod_{t=t_0}^{T_i} \prod_{k=0}^4 \Pr[d_{kit} = 1 | h_{it}, \mathbb{E}_{it}; \Lambda, \Theta]^{d_{kit}} \\ &\times \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_0}^{T_i} \prod_{j=1}^4 \Pr[y_{jit+1} | h_{it}, \tilde{\mu}_j; \Theta]^{d_{jit}} \right\} dF_{\mathcal{M}}(\tilde{\mathcal{M}}; \Delta_s) \end{aligned} \quad (\text{S25})$$

Equation (S25) 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 multidimensional integral.<sup>54</sup>

<sup>54</sup> 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 (S25).

Using (S25), the log likelihood can be separated additively:

$$\ln \mathcal{L}_i = \ln \mathcal{L}_i^d + \ln \mathcal{L}_i^y \quad (\text{S26})$$

The first stage of the estimation procedure utilizes the income term of the log likelihood to obtain estimates of  $\Theta$  and  $\Delta_s$ . These estimates are used in the second stage to estimate  $\Lambda$ . The scale of  $\Theta$ ,  $\Delta_s$ , and  $\rho$  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.

## C.2 First Stage Detailed

Since  $\ln \mathcal{L}_i$  is additively separable,  $\ln \mathcal{L}_i^y$  is used to consistently estimate  $\Theta$  and  $\Delta_s$  using an EM algorithm. 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

$$\ln \mathcal{L}_i^y(\mathcal{M}_i) = \sum_{t=t_{i0}}^{T_i} \sum_{j=0}^4 d_{jit} \ln \Pr [y_{jit+1} | h_{it}, \mu_j; \Theta] \quad (\text{S27})$$

Starting from a guess of parameters  $\langle \Theta^0, \Delta^0 \rangle$ , the EM algorithm iterates 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  $m$ th iteration:

$$E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i) | \cdot] \quad (\text{S28})$$

2. *Maximization Step.* Find the new iterated value of the vector of parameters by maximizing

the expression obtained in the expectation step:

$$\langle \Theta^{m+1}, \Delta^{m+1} \rangle = \max_{\langle \Theta, \Delta \rangle} \sum_i E_m [\ln \mathcal{L}_i^y(\mathcal{M}_i) | \cdot] \quad (\text{S29})$$

### C.2.1 Expectation Step

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

$$\mathbb{E}_i^m = \left( (\Delta_s^m)^{-1} + \Psi_i \right)^{-1} \mathbf{W}_i \quad (\text{S30})$$

$$\mathbb{V}_i^m = \left( (\Delta_s^m)^{-1} + \Psi_i \right)^{-1} \quad (\text{S31})$$

the  $k$ th component of the  $\mathbf{W}_i$  vector is

$$\mathbf{W}_{i\{k\}} = \frac{\sum_{t=1}^T d_{kit} (y_{kit} - h'_{it} \boldsymbol{\theta}_k)}{\sigma_{\eta_k}^{2,m}} \quad (\text{S32})$$

and the diagonal components of the square diagonal matrix  $\Psi_i$  are

$$\Psi_{i\{k,k\}} = \frac{\sum_{t=1}^T d_{kit}}{\sigma_{\eta_k}^{2,m}} \quad (\text{S33})$$

Given  $\mu_{ki}$  and the distribution of  $\eta_{kit}$

$$\begin{aligned} \log \Pr[y_{kit} | h_{it}, \boldsymbol{\mu}_k; \Theta] &= \log \left( \frac{1}{\sqrt{2\pi\sigma_{\eta_k}^2}} \exp \left\{ \frac{-(y_{kit} - h'_{it} \boldsymbol{\theta}_k - \mu_{ki})^2}{2\sigma_{\eta_k}^2} \right\} \right) \\ &= -\frac{1}{2} \log(2\pi\sigma_{\eta_k}^2) - \frac{1}{2\sigma_{\eta_k}^2} (y_{kit} - h'_{it} \boldsymbol{\theta}_k - \mu_{ki})^2 \end{aligned} \quad (\text{S34})$$

Therefore, the expectation step of the EM algorithm yields

$$\begin{aligned}
E_m [\log \mathcal{L}_i^y] &= - \sum_{t=1}^T \sum_{k=1}^4 d_{kit} \cdot E_m \left[ \frac{1}{2} \log (2\pi\sigma_{\eta_k}^2) + \frac{1}{2\sigma_{\eta_k}^2} (y_{kit} - h'_{it}\theta_k - \mu_{ki})^2 \right] \\
&= - \sum_{t=1}^T \sum_{k=1}^4 d_{kit} \left[ \frac{1}{2} \log (2\pi\sigma_{\eta_k}^2) + \frac{1}{2\sigma_{\eta_k}^2} \left( \mathbb{V}_{i\{k,k\}}^m + (y_{kit} - h'_{it}\theta_k - \mathbb{E}_{i\{k\}}^m)^2 \right) \right]
\end{aligned} \tag{S35}$$

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

### C.2.2 Maximization Step

Following the expectation step, the maximization step entails maximizing (S35) in order to obtain  $\Theta^{m+1}$ . In fact, each  $\theta_k^{m+1}$  is given by

$$\theta_k^{m+1} = \arg \min_{\theta_k} \sum_{i=1}^N \sum_{t=1}^T d_{kit} \left( y_{it} - h'_{it}\theta_k - \mathbb{E}_{i\{k\}}^m \right)^2 \tag{S36}$$

which yields

$$\theta_k^{m+1} = (H'W_kH)^{-1}H'W_kY_k \tag{S37}$$

where  $H$  is the  $[NT \times \#(\theta_k)]$  matrix that stacks together all values of  $h'_{it}$ ,  $Y_k$  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 (S35) and the estimated values of  $\theta_k^{m+1}$ ,  $\sigma_{\eta_k}^{2,m+1}$  has the closed form solution

$$\sigma_{\eta_k}^{2,m+1} = \frac{\sum_{i=1}^N \sum_{t=1}^T d_{kit} \left( \mathbb{V}_{i\{k,k\}}^m + \left( y_{it} - h'_{it}\theta_k^{m+1} - \mathbb{E}_{i\{k\}}^m \right)^2 \right)}{\sum_{i=1}^N \sum_{t=1}^T d_{kit}} \tag{S38}$$

Consistent with equation (S29), the maximization step also includes maximization of the expected value of the log likelihood of  $\mathcal{M}_i$  to update population parameter  $\Delta_s^{m+1}$ . This step is included in

the summary below.

### C.2.3 Combining the Steps

A summary of the EM algorithm is

↔ Step 1: Given  $m$ th iteration values  $\left\{ \theta_k^m, \sigma_{\eta_k}^{2,m} \right\}_{k \in \{1, \dots, 4\}}$  and  $\{\Delta_s^m\}_{s \in \{1, \dots, 4\}}$ , solve for  $\mathbb{E}_i^m$  and  $\mathbb{V}_i^m$  using (S30) and (S31).

↔ Step 2: Update population parameter  $\Delta_s^{m+1}$  for education level  $s$  as

$$\Delta_s^{m+1} = \frac{1}{N_s} \sum_{i=1}^N \sum_{s=1}^4 \delta_{is} (\mathbb{V}_i^m + \mathbb{E}_i^m \mathbb{E}_i^{m'}) \quad (\text{S39})$$

where  $\delta_{is}$  is an indicator of individual  $i$  having education level  $s$  and  $N_s = \sum_i \delta_{is}$ . Equation (S39) follows from maximization of the expected value of the log likelihood of  $\mathcal{M}_i$ ,  $E_m[\log f_{\mathcal{M}}(\mathcal{M}_i)]$  (Anderson and Olkin, 1985).

↔ Step 3: For each occupation  $k > 0$ , new iteration values  $\theta_k^{m+1}$  are obtained using equation (S37) and new iteration values  $\sigma_{\eta_k}^{2,m+1}$  are obtained using equation (S38).

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

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

where

$$\tilde{\mathcal{L}}_i^{y,m} = \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_{i0}}^{T_i} \prod_{j=1}^4 \Pr [y_{jit+1} | h_{it}, \tilde{\mu}_j; \Theta^m]^{d_{jit}} \right\} f_{\mathcal{M}}(\tilde{\mathcal{M}}; \Delta_s^m) \quad (\text{S40})$$

is computed using Monte Carlo integration.  $\varepsilon$  is set to be  $1 \times 10^{-4}$ .

Consistent estimates of individual beliefs are also obtained for use in the second stage. This computation uses the point estimates for  $\Theta$  and  $\Delta_s$ , the history of signals received by every indi-

vidual, Bayes' Rule, and the rational expectations assumption regarding the individual's prior.

### C.3 Second Stage Detailed

In order to estimate  $\Lambda$ , the second stage follows Hotz and Miller (1993) and takes advantage of the expression derived in Proposition 2 that maps 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<sup>55</sup>

$$p_{kit}(h_{it}, \mathbb{E}_{it}) = \frac{\exp(V_k(h_{it}, \mathbb{E}_{it}))}{1 + \sum_{k' > 0} \exp(V_{k'}(h_{it}, \mathbb{E}_{it}))} \quad (\text{S41})$$

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

$$V_k(h_{it}, \mathbb{E}_{it}) = -\ln \alpha_{kit}(h_{it}) - (b_{\tau(t)} - 1) \ln E_t \left[ v_{kit+1} \prod_{s=1}^{T-t} \left( \frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{0it}^{(s)})} \right)^{\phi_t(s)} \middle| \mathbb{E}_{it}, h_{it} \right] \quad (\text{S42})$$

An iterative algorithm is implemented that maximizes the log likelihood of the data while searching over the space of parameters and ccps. The algorithm is initialized with flexible parametric versions of the future conditional choice probabilities estimated from the data (Figure S2), where the beliefs, estimated in the first stage, are also treated as data. It entails the following two steps:

1. *Maximization Step.* Plug the estimated ccps in equation (S41) and maximize the log likelihood of the observed choices.
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. In the model, individuals have perfect foresight about their marital status. However, their entire marital status vector up

---

<sup>55</sup> Because the posterior variance is redundant conditional on the vector  $h_{it}$  (Appendix B above), in this section the posterior mean  $\mathbb{E}_{it}$  substitutes the individual beliefs  $\mathbb{B}_{it}$  in the individual state. Moreover, although the bond prices  $b_{\tau(t)}$  are part of the state, to save on notation they are excluded from the arguments of  $p_{kit}(\cdot)$ . This is correctly handled in estimation.

to period  $T$  is not always observed. When necessary, 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). Effectively it amounts to individuals getting married at age 23 and remaining married until age 50.

### C.3.1 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

$$\frac{1}{NT} \sum_i \sum_t \sum_{k=0}^4 d_{kit} \ln p_{kit}(h_t, \mathbb{E}_t) \quad (\text{S43})$$

The expectation in the expression for  $V_k(h_t, \mathbb{E}_t)$  in equation (S42) can be written as

$$\begin{aligned} & E_t \left[ v_{kit+1} \prod_{s=1}^{T-t} \left( \frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{0it}^{(s)})} \right)^{\phi^{(s)}} \middle| \mathbb{E}_{it}, h_{it} \right] \\ &= \int_{\zeta_k} \left\{ \exp \left( \frac{-\rho \bar{L}_k y_{kit+1}(h_{it})}{b_{\tau(t+1)}} \right) \prod_{s=1}^{T-t} \left( \frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi^{(s)}} \right\} dF_{\zeta_k}(\zeta_k | \mathbb{E}_{it}, h_{it}) \\ &= \int_{\zeta_k} \left\{ \exp \left( \frac{-\rho \bar{L}_k(f_k(h_{it}, \omega_i; \theta_k) + \zeta_k)}{b_{\tau(t+1)}} \right) \prod_{s=1}^{T-t} \left( \frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi^{(s)}} \right\} dF_{\zeta_k}(\zeta_k | \mathbb{E}_{it}, h_{it}) \end{aligned} \quad (\text{S44})$$

I compute the value of (S44) using Monte Carlo integration.

Given a value for  $\rho$  the model becomes a simple logit in the  $\alpha$  parameters and the Monte Carlo integral is<sup>56</sup>

$$B_{kit}(\rho) = \frac{1}{S} \sum_s \left\{ \exp \left( \frac{-\rho \bar{L}_k(f_k(h_{it}; \theta_k) + \zeta_k^s)}{b_{\tau(t+1)}} \right) A_{kit}(\zeta_k^s) \right\} \quad (\text{S45})$$

---

<sup>56</sup> Recall that the scale of parameters  $\Theta$ ,  $\Delta_s$ , and  $\rho$  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

$$A_{kit}(\zeta_k^s) = \prod_{s=1}^{T-t} \left( \frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k^s))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi^{(s)}} \quad (\text{S46})$$

varies across the signals  $\zeta_k^s$  drawn for integration. The draws come from the distribution of signals conditional on current beliefs

$$\zeta_k^s = \mu_{ki} + \eta_{kit} \sim N(\mathbb{E}_{\{k\}it}, \mathbb{V}_{\{k,k\}it} + \sigma_{\eta_k}^2) \quad (\text{S47})$$

Using the definition of  $\alpha_{kt}(h_t)$  in equation (5), equation (S42) can be rewritten as

$$V_k(h_t, \mathbb{E}_t) = -h'_{it} \alpha_k - C_{kit}(\rho) \quad (\text{S48})$$

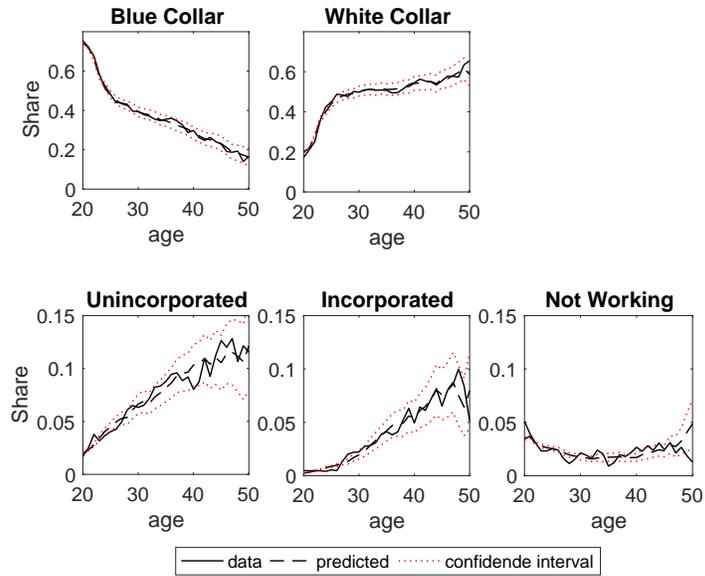
where

$$C_{kit}(\rho) = (b_{\tau(t)} - 1) \ln B_{kit}(\rho) \quad (\text{S49})$$

Equation (S48) is then substituted into (S41). In estimation, the log likelihood is maximized conditional on a value of  $\rho$ . Search over values for  $\rho$  is then undertaken and the value that maximizes the log likelihood is selected. This procedure is faster than searching over all the parameter space at once because it avoids computing the Monte Carlo integral in (S44) more than once for each value of  $\rho$ .

### C.3.2 CCP Step

For a given value of utility parameters the model is solved backwards (Appendix D below) and new model-generated ccps are obtained. This new ccps are fed into the maximization step and new utility parameters are obtained. 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 iterations 4 and 5 is 9.4. Since the first search was initialized from 10 different initial points the solution is unlikely to be local. Notice that the estimated parameters at each iteration are consistent because the ccps that initialize the process are themselves consistent. In particular, Figure S2 shows that the initial ccps fed into the iterative process fit the data well.



**APPENDIX FIGURE S2: Average Conditional Choice Probabilities**

Notes: Average estimated conditional choice probabilities (in first iteration) against the share of people choosing the alternative. Dashed lines represent 95% confidence intervals around the predicted ccps.

## D Results Appendix

### D.1 Parameter Estimates

**APPENDIX TABLE S2: Income Parameters**

	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
<i>Constant</i>	0.927	(0.008)	0.847	(0.021)	1.304	(0.050)	0.730	(0.159)
<i>Black</i>	-0.204	(0.009)	-0.122	(0.015)	-0.180	(0.043)	-0.097	(0.106)
<i>Some College</i>	0.222	(0.009)	0.129	(0.016)	-0.095	(0.044)	0.597	(0.192)
<i>College</i>	0.323	(0.022)	0.456	(0.021)	0.537	(0.109)	0.411	(0.140)
<i>More than College</i>	0.251	(0.026)	0.624	(0.022)	0.934	(0.122)	1.842	(0.325)
<i>Married</i>	0.053	(0.008)	0.213	(0.014)	0.019	(0.047)	0.718	(0.132)
$1\{exp_{bc} \geq 1\}$	0.081	(0.005)	-0.077	(0.014)	-0.233	(0.054)		
$1\{exp_{bc} \geq 2\}$	0.074	(0.006)					-0.268	(0.161)
$1\{exp_{bc} \geq 3\}$	0.041	(0.007)						
$1\{exp_{bc} \geq 4\}$	0.056	(0.006)						
$1\{exp_{bc} \geq 5\}$	0.067	(0.007)					-0.508	(0.162)
$1\{exp_{bc} \geq 6\}$	0.142	(0.007)						
$1\{exp_{bc} \geq 7\}$			0.119	(0.022)				
$1\{exp_{wc} \geq 1\}$	0.133	(0.010)	0.126	(0.006)				
$1\{exp_{wc} \geq 2\}$			0.127	(0.006)	0.019	(0.059)	-0.765	(0.154)
$1\{exp_{wc} \geq 3\}$			0.192	(0.011)	0.044	(0.066)	0.824	(0.156)
$1\{exp_{wc} \geq 5\}$			0.140	(0.011)				
$1\{exp_{wc} \geq 6\}$	0.209	(0.026)			0.480	(0.063)		
$1\{exp_{wc} \geq 7\}$			0.178	(0.012)				
$1\{exp_{wc} \geq 8\}$			0.410	(0.015)			0.627	(0.163)
$1\{exp_{eu} \geq 1\}$			-0.177	(0.035)	0.219	(0.031)		
$1\{exp_{eu} \geq 3\}$							0.229	(0.151)
$1\{exp_{eu} \geq 4\}$			0.743	(0.177)	0.407	(0.043)		
$1\{exp_{eu} \geq 7\}$					0.280	(0.048)		
$1\{exp_{ei} \geq 1\}$					-0.451	(0.078)	0.433	(0.110)
$1\{exp_{ei} \geq 2\}$			1.004	(0.156)				
$1\{exp_{ei} \geq 3\}$					1.867	(0.430)		
$1\{exp_{ei} \geq 5\}$							1.078	(0.148)
$1\{exp_e \geq 1\}$	-0.022	(0.017)						
$1\{exp_e \geq 2\}$	-0.156	(0.026)						
$1\{exp_e \geq 5\}$	0.365	(0.076)						
<i>Obs</i>	8902		9957		1403		602	

Notes: Hourly income measured in \$10s. Standard errors corrected for 2-stage estimation using subsampling estimation over 100 subsamples without replacement. Estimated parameters of equation (2). 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, specially in entrepreneurial occupations. Steps were chosen using statistical significance in a preliminary OLS regression as a baseline. 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.

**APPENDIX TABLE S3: Population Ability Covariance Matrices**

	High School							
	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Blue Collar	0.15	(0.005)						
White Collar	0.13	(0.005)	0.11	(0.008)				
Unincorporated	0.20	(0.010)	0.17	(0.013)	0.27	(0.028)		
Incorporated	0.04	(0.017)	0.03	(0.014)	0.05	(0.023)	0.04	(0.008)

	Some College							
	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Blue Collar	0.23	(0.009)						
White Collar	0.26	(0.015)	0.32	(0.027)				
Unincorporated	0.16	(0.024)	0.14	(0.029)	0.28	(0.068)		
Incorporated	0.45	(0.045)	0.77	(0.089)	0.35	(0.078)	4.60	(1.147)

	College							
	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Blue Collar	0.45	(0.074)						
White Collar	0.33	(0.041)	0.57	(0.07)				
Unincorporated	0.29	(0.100)	0.38	(0.083)	3.52	(0.710)		
Incorporated	0.71	(0.073)	0.85	(0.146)	-0.11	(0.286)	1.66	(0.209)

	More than College							
	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Blue Collar	0.37	(0.044)						
White Collar	0.22	(0.044)	0.87	(0.075)				
Unincorporated	-0.41	(0.128)	0.66	(0.143)	3.03	(0.297)		
Incorporated	-0.29	(0.650)	1.82	(0.125)	2.35	(0.885)	10.88	(1.690)

Notes: Standard errors corrected for 2-stage estimation using subsampling estimation over 100 subsamples without replacement. Covariance matrix of the joint distribution of unobserved ability conditional on education, denoted  $\Delta_s$ .

**APPENDIX TABLE S4: Idiosyncratic Variance**

Blue Collar		White Collar		Unincorporated		Incorporated	
coeff	se	coeff	se	coeff	se	coeff	se
0.30	(0.012)	0.96	(0.063)	2.47	(0.180)	8.00	(1.089)

Notes: Standard errors corrected for 2-stage estimation using subsampling estimation over 100 subsamples without replacement. Idiosyncratic hourly income variance in every occupation, denoted  $\sigma_{\eta_k}^2$ .

**APPENDIX TABLE S5: Utility Parameters**

$\rho$		coeff	se						
		0.040	(0.0023)						
High School									
$\alpha$		Blue Collar		White Collar		Unincorporated		Incorporated	
		coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	<i>constant</i>	-1.818	(0.124)	-0.965	(0.164)	-0.588	(0.182)	-1.551	(0.457)
	<i>black</i>	0.730	(0.083)	1.182	(0.116)	0.884	(0.157)	1.659	(0.578)
	<i>married</i>	-0.689	(0.065)	-0.369	(0.087)	-0.544	(0.102)	0.072	(0.446)
Entry Cost	<i>constant</i>	-4.778	(1.699)	2.967	(0.616)	5.347	(0.801)	12.963	(0.251)
	<i>age/10</i>	2.846	(0.800)	-0.132	(0.194)	-0.549	(0.232)	-2.168	(0.084)
	$\omega_i/10^3$	2.775	(5.833)	-1.529	(0.856)	0.451	(2.270)	-0.246	(0.196)
	$(age/10) \cdot (\omega_i/10^3)$	-1.145	(2.777)	0.413	(0.299)	-0.224	(0.698)	0.060	(0.067)
Some College									
$\alpha$		Blue Collar		White Collar		Unincorporated		Incorporated	
		coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	<i>constant</i>	-1.985	(0.118)	-1.933	(0.144)	-1.745	(0.182)	-1.258	(0.334)
	<i>black</i>	0.452	(0.121)	0.839	(0.139)	1.067	(0.231)	1.067	(0.320)
	<i>married</i>	-0.933	(0.101)	-0.787	(0.108)	-0.673	(0.125)	-0.732	(0.253)
Entry Cost	<i>constant</i>	-3.842	(0.419)	3.647	(0.579)	4.978	(0.407)	11.285	(3.445)
	<i>age/10</i>	2.057	(0.171)	-0.078	(0.164)	-0.329	(0.109)	-1.670	(0.774)
	$\omega_i/10^3$	-0.453	(1.081)	-2.954	(0.566)	-0.020	(0.166)	-0.497	(0.408)
	$(age/10) \cdot (\omega_i/10^3)$	0.211	(0.496)	1.082	(0.231)	-0.018	(0.068)	0.166	(0.228)
College									
$\alpha$		Blue Collar		White Collar		Unincorporated		Incorporated	
		coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	<i>constant</i>	-2.650	(0.181)	-3.614	(0.194)	-3.012	(0.258)	-4.462	(0.337)
	<i>black</i>	0.184	(0.237)	1.437	(0.298)	1.779	(0.342)	15.975	(3.110)
	<i>married</i>	-0.210	(0.156)	0.374	(0.184)	0.088	(0.203)	1.078	(0.210)
Entry Cost	<i>constant</i>	-3.227	(0.810)	1.147	(0.560)	5.707	(2.343)	7.775	(1.233)
	<i>age/10</i>	1.929	(0.322)	0.815	(0.196)	-0.503	(0.525)	-0.802	(0.288)
	$\omega_i/10^3$	-0.821	(1.519)	-1.560	(1.086)	-3.238	(0.838)	-2.799	(0.906)
	$(age/10) \cdot (\omega_i/10^3)$	0.592	(0.615)	0.196	(0.421)	0.964	(0.288)	0.750	(0.274)
More than College									
$\alpha$		Blue Collar		White Collar		Unincorporated		Incorporated	
		coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	<i>constant</i>	-1.075	(0.190)	-1.994	(0.214)	-1.462	(0.271)	-1.005	(0.403)
	<i>black</i>	-0.674	(0.207)	-0.295	(0.189)	0.470	(0.663)	-0.561	(0.490)
	<i>married</i>	-0.387	(0.147)	-0.316	(0.145)	-0.713	(0.185)	-0.725	(0.267)
Entry Cost	<i>constant</i>	-5.807	(0.606)	-3.242	(0.788)	5.258	(0.619)	9.303	(0.796)
	<i>age/10</i>	2.655	(0.221)	2.294	(0.321)	-0.228	(0.160)	-0.930	(0.195)
	$\omega_i/10^3$	2.128	(1.010)	0.955	(1.482)	-0.510	(0.372)	-3.819	(0.557)
	$(age/10) \cdot (\omega_i/10^3)$	-0.586	(0.369)	-0.589	(0.572)	0.055	(0.096)	1.076	(0.195)

Notes: Standard errors corrected for 2-stage estimation using subsampling estimation over 100 subsamples without replacement.  $\omega_i$  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 (4) and (5).

## D.2 Solving the Model

As mentioned in Appendix C, solution of the model is required in order to provide new estimates of the conditional choice probabilities at every iteration of the second stage in the estimation process. The model is solved using the same representation obtained in Proposition 2 and summarized in equation (S42). 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 age-specific 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 (S42) and the future choice paths described in Proposition 2.<sup>57</sup>

↔ *Step 2.* Obtain relevant ccps for period  $t$  using equation (S41).

↔ *Step 3.* Obtain parametric versions, characterized by  $\Omega_t$ , of the ccps for period  $t$ . 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,  $p_{0it}(h_{it}, \mathbb{E}_{it})$ , and a the following parameterization

$$\exp(X_{it}'\Omega_t)/(1 + \exp(X_{it}'\Omega_t)) \quad (\text{S50})$$

where  $X_{it}$  includes multiple interactions of components of the state.

↔ *Step 4.* If  $t = t_0$ , stop. Otherwise, set  $t = t - 1$ , go back to Step 1, and use the collection of parametric ccps obtained so far ( $\{\Omega_r\}_{r=t}^T$ ) for the representation of the continuation value.

This algorithm yields a collection of future ccps,  $\{\Omega_t\}_{t=t_0}^T$ , that characterize the value function at any period  $t$  using the representation obtained in Proposition 2.

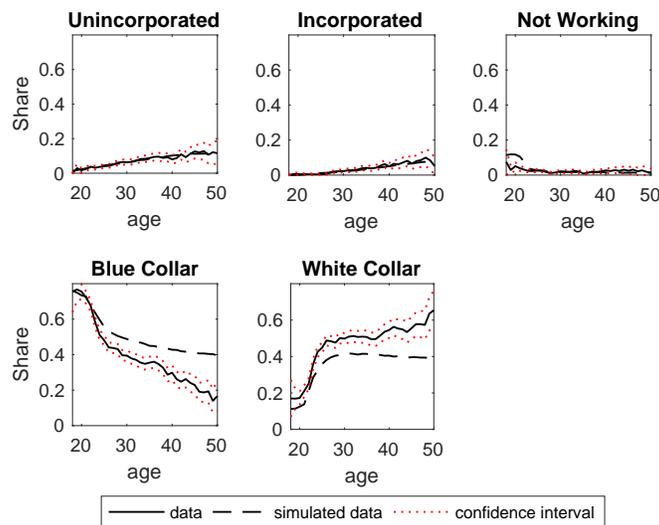
---

<sup>57</sup> 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.

### D.3 Model Fit

In order to extend the assessment of goodness of fit, an initial state is generated and the model is simulated forward using the collection of future ccps implied by the model,  $\{\Omega_t\}_{t=t_0}^T$  in equation (S50), that characterize the value function using the representation obtained in Proposition 2. For comparison against the data, initial states are obtained by drawing from the data. First, a collection of initial states formed by race, education, entry age, year of entry, and permanent wealth is drawn from the initial states observed in the data. 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 Appendix C. Third, ability for each individual drawn from the data is set at 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.

Comparing simulated data against the sample, Figure S3 shows that the model fits participation shares and transitions very well for incorporated and unincorporated entrepreneurship over the life cycle, and less precisely for white collar and blue collar.<sup>58</sup>



**APPENDIX FIGURE S3: Simulated versus Observed Choice Rates**

Notes: Dotted line indicates a 95% confidence interval.

<sup>58</sup> Figure S3 is different from Figure 4 since the former, stricter measure of fit, simulates forward entire paths from an initial state and the latter takes the state from the data at every age as given and simulates current choices.

First-entry statistics presented in Table S6 show that the model replicates the absence of young entrepreneurs. The model captures well the proportion of individuals who attempt entrepreneurial occupations by age 40, and it 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 data, simulated individuals attempting entrepreneurship for the first time tend to have more prior experience. Moreover, consistent with the similarities between white collar work and incorporated entrepreneurship highlighted in Section 2, first-time unincorporated entrepreneurs tend to have more blue collar experience, whereas the opposite is true for first-time incorporated entrepreneurs.

**APPENDIX TABLE S6:** First Entry: Observed and Simulated

<i>Data</i>				
	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
<i>exp<sub>bc</sub></i>	-	1.99	3.48	2.05
<i>exp<sub>wc</sub></i>	1.08	-	3.37	6.58
<i>exp<sub>eu</sub></i>	0.03	0.15	-	0.96
<i>exp<sub>ei</sub></i>	0.01	0.02	0.22	-

<i>Model</i>				
	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
<i>exp<sub>bc</sub></i>	-	3.32	4.51	3.73
<i>exp<sub>wc</sub></i>	0.58	-	3.17	5.53
<i>exp<sub>eu</sub></i>	0.08	0.25	-	0.48
<i>exp<sub>ei</sub></i>	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 S7 compares transition matrices from the data and simulated. In the model, occupations are less absorbing than in the data. However, consistent 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 S8 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.

**APPENDIX TABLE S7:** Transition Patterns: Observed and Simulated

<i>Data</i>					
	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
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
Unemployed	0.37	0.15	0.03	0.00	0.44

<i>Model</i>					
	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
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
Unemployed	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  $t + 1$ .

**APPENDIX TABLE S8:** Spells: Observed and Simulated

<i>Data</i>						
	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

<i>Model</i>						
	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.

Table S9 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.

**APPENDIX TABLE S9:** Income: Observed and Simulated

<i>Data</i>				
	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

<i>Model</i>				
	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

## D.4 Certainty Equivalent

### D.4.1 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 occupation  $k$  at age  $t$  is

$$\bar{y}_{kt+1} = f_k(h_{it}; \theta_k) + \mathbb{E}_t\{y_k\} \quad (\text{S51})$$

and he considers the variance of his hourly income to be

$$\sigma_{kt}^2 = \mathbb{V}_t\{y_k\} + \sigma_{\eta_k}^2 \quad (\text{S52})$$

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

$$-\exp\{-\rho \bar{L}_k y_k^c\} = -\exp\left\{-\rho \bar{L}_k \bar{y}_{kt+1} + \frac{\rho^2 \bar{L}_k^2}{2} \sigma_{kt}^2\right\} \quad (\text{S53})$$

which yields

$$y_k^c = \bar{y}_{kt+1} - \frac{\rho \bar{L}_k}{2} \sigma_{kt}^2 \quad (\text{S54})$$

Recall, in estimation  $\bar{L}_k$  is substituted with  $\bar{L}_k/100$ .

#### D.4.2 Dynamic

In order to obtain the dynamic version of the certainty equivalent use equation (S14) to find the quantity  $y_k^c$  such that

$$\begin{aligned} & E_t \left[ A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp \left( \frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}} \right) \middle| \mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}} \\ &= E_t \left[ A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t) \exp \left( \frac{-\rho \bar{L}_k y_k^c}{b_{\tau(t+1)}} \right) \middle| \mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}} \end{aligned} \quad (\text{S55})$$

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

$$y_k^c = - \left( \frac{b_{\tau(t+1)}}{\rho \bar{L}_k} \right) \ln \left( \frac{E_t [A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1} | \mathbb{E}_t, h_t]}{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t)} \right) \quad (\text{S56})$$

#### D.5 Monetary Value of Entry Costs

The monetary value of entry costs is obtained using equation (S14). From equation (5) one can separate the non-pecuniary costs in two factors, one corresponding to the entry costs,  $\alpha_{kt}^e(h_t)$  and the other corresponding to all other non-pecuniary costs. Hence, let  $\alpha_{kt}(h_t) = \alpha_{kt}^o(h_t) \times \alpha_{kt}^e(h_t)$ . Next, use equation (5) to find 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  $\psi$  that must be given to the individual to leave him indifferent between (a) receiving  $\psi$  and facing entry costs and (b) not receiving  $\psi$  but facing no entry costs. The quantity  $\psi$  solves:

$$\begin{aligned}
& \alpha_{kt}^e(h_t)^{1/b_{\tau(t)}} E_t \left[ A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right) \middle| \mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}} \\
& = E_t \left[ A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho(\bar{L}_k y_{kt+1}(h_t) - \psi)}{b_{\tau(t+1)}}\right) \middle| \mathbb{E}_t, h_t \right]^{1-1/b_{\tau(t)}}
\end{aligned} \tag{S57}$$

which yields

$$\psi = \frac{\ln \alpha_{kt}^e(h_t)}{\rho} \frac{b_{\tau(t+1)}}{b_{\tau(t)} - 1} \tag{S58}$$

Recall that the quantity  $\bar{L}_k y_{kt+1}$  is written in thousands of dollars in estimation. Hence, the value of  $\psi$  is in thousands of dollars.

## D.6 Alternative Regimes

In order to increase precision and facilitate comparison across options, in this section ability is not approximated from the data using Bayes' rule. Instead, individuals' ability vectors are drawn from the estimated distributions in Table S3. 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.<sup>59</sup> Marital status paths follow the same restriction specified in Appendix C. The initial state and bond price sequence used in the decomposition are also used for policy counterfactuals.

These counterfactual regimes are described below. Extended results are presented in Table S11.

↔ 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. This average is computed using the returns to experience

---

<sup>59</sup> 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.

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 (Figure 6a). Then, the fixed hourly return to individuals in occupation  $k$  under this counterfactual is

$$\bar{y}_k = \sum_{x=0}^{20} R_k(x) \quad (\text{S59})$$

This exercise yields the following values:

**APPENDIX TABLE S10:** Average Income for No Learning-by-doing Counterfactual

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

Notes: Computed using equation (S59) and the profiles in Table S2.

Individuals under this counterfactual continue to have different 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 addition, 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 (Table S4) plus the value of the ability variance (Table S3). In terms of equation (2), this amounts to changing the value of the idiosyncratic income variance in occupation  $k$  from just  $\sigma_{\eta_k}$  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  $k' \neq k$  (Figure 6b) are set to be zero.
- ↔ **C4: *Uncorrelated learning (no cross-occupation learning about ability).*** In this counterfactual, individuals use an alternative prior variance-covariance matrix to update beliefs. This alternative prior variance-covariance matrix results from diagonalizing 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 resulting from the idiosyncratic income variation. In other words, this counterfactual is the same as counterfactual C2 plus setting the idiosyncratic income 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 (Table S5).
- ↔ *C7 and C8: Permanent wealth at median and 99th percentile.* In these counterfactuals, all individuals are given the median and 99th percentile value of permanent wealth, respectively.

**APPENDIX TABLE S11: Comparison of Counterfactual Regimes**

<i>Unincorporated</i>									
	Baseline	C1	C2	C3	C4	C5	C6	C7	C8
Ever tried	0.31	0.39	0.34	0.28	0.32	0.35	0.54	0.31	0.44
Ever tried in first 5 years	0.08	0.08	0.11	0.06	0.08	0.12	0.31	0.08	0.11
PVI if ever tried	518	510	663	493	481	666	618	514	536
Spell duration	2.17	2.16	2.81	2.37	2.07	2.89	3.10	2.15	1.91
Participation rate at age 40	0.10	0.13	0.14	0.10	0.10	0.15	0.26	0.10	0.13
<b>At first entry</b>									
Ability (10\$ per hour)	0.05	0.05	0.59	0.09	-0.09	0.57	0.03	0.05	0.02
Belief (10\$ per hour)	0.04	0.04	-	0.07	0.00	-	0.01	0.04	-0.02
Age	34.07	34.88	32.80	34.91	33.99	32.53	28.04	34.24	34.30
$exp_{bc}$	6.76	8.37	6.01	9.76	6.47	5.90	0.85	6.80	5.40
$exp_{wc}$	4.77	3.86	4.25	2.55	5.09	4.04	3.88	4.92	6.67
<b>Overall</b>									
Ability (10\$ per hour)	0.37	0.39	1.16	0.45	0.21	1.12	0.23	0.37	0.36
College or more	0.55	0.58	0.65	0.63	0.49	0.65	0.44	0.53	0.52

<i>Incorporated</i>									
	Baseline	C1	C2	C3	C4	C5	C6	C7	C8
Ever tried	0.15	0.29	0.20	0.14	0.13	0.26	0.26	0.16	0.20
Ever tried in first 5 years	0.02	0.02	0.05	0.01	0.02	0.07	0.12	0.02	0.06
PVI if ever tried	757	624	1115	622	576	1100	840	749	936
Spell duration	2.88	2.98	5.11	2.91	2.62	5.57	3.39	2.84	3.09
Participation rate at age 40	0.04	0.09	0.10	0.04	0.03	0.14	0.13	0.04	0.08
<b>At first entry</b>									
Ability (10\$ per hour)	0.55	0.43	1.73	0.50	-0.25	1.52	0.20	0.61	0.50
Belief (10\$ per hour)	0.64	0.49	-	0.54	0.00	-	0.26	0.71	0.59
Age	38.62	40.52	36.06	40.59	38.74	35.43	29.19	39.44	33.88
$exp_{bc}$	6.95	11.43	5.98	13.50	7.03	6.07	0.45	7.09	2.70
$exp_{wc}$	8.14	5.68	6.52	3.97	8.09	5.85	5.18	8.75	7.61
<b>Overall</b>									
Ability (10\$ per hour)	1.18	0.92	2.70	1.12	0.24	2.37	0.42	1.18	1.17
College or more	0.70	0.62	0.66	0.64	0.61	0.63	0.37	0.70	0.82

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 about ability. Both ability and the mean of the belief are in 10\$ per hour.  $exp_{bc}$  and  $exp_{wc}$  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 C1 to C6 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 income variance is set to be  $\sigma_{\eta_k} + V_{\{k,k\}}$ . **C3** sets the cross-occupation returns to experience to zero. **C4** shuts down correlated learning. **C5** is the full information model without uncertainty. In this counterfactual, the idiosyncratic income variance  $\sigma_{\eta_k}$  is set to 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 sample of simulated individuals, including their ability vector, is kept constant across all counterfactual regimes. **C7** and **C8** set everyone's permanent wealth at the median and at the 99th percentile, respectively.

## D.7 Calibrating the Quality of an Entrepreneurship Education Program

In a fashion that resembles the learning about ability mechanism of the model developed here, von Graevenitz et al. (2010) characterize entrepreneurship education as a source of information. They measure students' beliefs about entrepreneurial aptitude before and after a mandatory entrepreneurship course that every business administration undergraduate student at a major German university must take. Using their model the authors predict, among other things, that the variance of beliefs across individuals (aggregate variance of beliefs) should increase as a result of the program. In Table 8 in their paper they find that this is actually the case: the aggregate standard deviation of beliefs in their most restricted sample increases by 8.24%. I use this finding to calibrate a value of information quality in my entrepreneurship education counterfactual that would generate the same increase in the aggregate variance of beliefs.

Since there is no initial dispersion in beliefs in my model, I assume that the initial variation in aggregate beliefs in von Graevenitz et al. (2010), which the authors associate to information received in life before entering the market, amounts to the variation generated by one signal of ability in my model. Moreover, since their program is at the college level, I take the distribution of completed education conditional on having more than high school from my data to generate the mapping. The entrepreneurship education program in Section 7 provides all young individuals with a signal about their incorporated entrepreneurial ability. Hence, I construct a mapping from information quality of a program into changes in the aggregate variance of beliefs about incorporated entrepreneurial ability. This mapping yields from varying the information quality of the program—characterized by  $s$  in equation (13)—in discrete steps and measuring the corresponding change in the aggregate variance of beliefs relative to the initial aggregate variance of beliefs. Using this mapping I locally interpolate to obtain a value of  $s$  that generates a change in aggregate beliefs of 8.24%. This exercise indicates that the information from the program is a little under half the quality of the information in the market ( $s = 2.1$ ).

## Appendix References

- ANDERSON, T. AND I. OLKIN, “Maximum-likelihood Estimation of the Parameters of a Multivariate Normal Distribution,” *Linear Algebra and its Applications* 70 (1985), 147 – 171.
- DEGROOT, M. H., *Optimal Statistical Decisions* (McGraw Hill, 1970).
- GAYLE, G.-L., L. GOLAN AND R. A. MILLER, “Promotion Turnover and the Compensation in the Executive Labor Market,” *Econometrica* 83 (2015), 2293–2369.
- GAYLE, G.-L. AND R. A. MILLER, “Has Moral Hazard Become a More Important Factor in Managerial Compensation?,” *American Economic Review* 99 (2009), 1740–1769.
- HOTZ, V. J. AND R. A. MILLER, “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *The Review of Economic Studies* 60 (1993), 497–529.
- JAMES, J., “Ability Matching and Occupational Choice,” SSRN Scholarly Paper ID 1937879, 2011.
- KREIDER, R. M. AND R. ELLIS, “Number, Timing, and Duration of Marriages and Divorces: 2009,” Technical Report, U.S. Census Bureau., 2011.
- MARGIOTTA, M. M. AND R. A. MILLER, “Managerial Compensation and the Cost of Moral Hazard,” *International Economic Review* 41 (2000), 418–438.
- RYAN, C. AND J. SIEBENS, “Educational Attainment in the United States: 2009,” Technical Report, U.S. Census Bureau., 2012.