INTERNATIONAL ECONOMIC REVIEW
Vol. 61, No. 2, May 2020 DOI: 10.1111/iere.12436

ENTREPRENEURSHIP OVER THE LIFE CYCLE: WHERE ARE THE YOUNG ENTREPRENEURS?*

BY ANDRÉS HINCAPIÉ[1]

Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A.

Most individuals do not start a business and, if they do, they start well into their 30s. To explain these stylized facts, I estimate a dynamic Roy model with experience accumulation, risk aversion, and imperfect information about ability using the Panel Study of Income Dynamics. Information frictions and income risk reduce entrepreneurship by up to 40% and 35%, respectively. Entry costs and information frictions explain most of the delayed entry. Results from counterfactual policies targeting delayed entry suggest that entrepreneurship education can yield higher returns than subsidies. Fostering young entrepreneurship yields higher returns than fostering old entrepreneurship.

1. INTRODUCTION

Less than 35% of individuals start a business during their careers and they tend to do so well into their 30s (Table 1). Although our understanding of the relative importance of the economic forces explaining these facts is limited, policymakers 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.[2] Without properly understanding the mechanisms explaining 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, nonpecuniary 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

*Manuscript received October 2018; revised February 2019.

[1] I thank my dissertation advisors Barton Hamilton, George-Levi Gayle, Robert Miller, and Limor Golan for their helpful comments and invaluable advice. I thank my editor, Holger Sieg, and three anonymous referees for greatly improving this article. 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. Please address correspondence to: Andrés Hincapié, Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. E-mail: *andres.hincapie@unc.edu*.

[2] The popular press has often shared the interest in young entrepreneurship (e.g., in the *30 under 30* collection by *Forbes*).

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 who 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.

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, 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, nonpecuniary motivations such as "being one's own boss" could also provide incentives for entry (Hamilton, 2000; Hurst and Pugsley, 2017).

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-30s, 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.[3] Moreover,

[3] 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, inclined to remain employed based on their success, have even less incentives to switch because entrepreneurship becomes more risky.

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.[4] 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.[5] I endogenize selection in a rational expectations framework 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).[6]

Estimates indicate that entrepreneurial ability displays higher variation than paid-employment ability, which both encourages individuals to attempt entrepreneurship 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 five years of white collar experience.

The deterring effect of risk aversion is nonnegligible. Using the structure of the model, a decomposition exercise indicates that eliminating income risk 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, that is, 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. Learning-by-doing and entry costs have stronger effects on the participation margin.

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 capture barriers to entrepreneurship not explicitly modeled. Short-run difficulties that young individuals with weaker credit histories and less savings often face when attempting entrepreneurship are captured with an age profile in the entry cost. Additionally, long-run advantages in access to resources reducing entry costs to entrepreneurship are captured through a permanent wealth measure (estimated as a fixed effect outside of the structural model). Younger individuals as well as individuals with lower permanent wealth face higher

[4] I abstract from other mechanisms such as personality traits (Humphries, 2018; Hamilton et al., 2019) and parental influence (Lindquist et al., 2015).

[5] Since individuals' beliefs change over time as they acquire information, the unobserved component cannot be controlled for using fixed effects (Gibbons et al., 2005).

[6] This implies that the representation must be updated with ccps that capture the nature of the new regime.

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.[7] The first one introduces a blanket subsidy for young incorporated entrepreneurship. The second one extends a strand of the literature that, although silent in terms of long-term outcomes, shows that entrepreneurship education, as a source of information, shifts individual beliefs (Oosterbeek et al., 2010; von Graevenitz et al., 2010). In particular, I provide a mapping from movements in beliefs, generated by entrepreneurship education of a given quality, into career choices and long-term outcomes. Both programs increase young entrepreneurship. A $25,000 subsidy has a return of 11.2 dollars in present value of income (PVI) per dollar of subsidy cost. To compare the returns of entrepreneurship education against those of subsidies, I use the structure of the model to calibrate the information quality of the entrepreneurship program studied by von Graevenitz et al. (2010), provide bounds for entrepreneurship education costs, and compare these bounds against average undergraduate, graduate, and MBA course costs using information from the National Center for Education Statistics and *U.S. News*. I find that entrepreneurship education programs tend to increase young entrepreneurship more than subsidies and can have higher returns. Additionally, I simulate the introduction of the policies at older ages and show that the returns of both policies decrease monotonically with age of implementation, suggesting that policymakers fostering entrepreneurship are right to focus on the young.

In the literature, my article 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.[8] I find that income risk aversion plays 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—which are difficult to justify empirically.[9] Third, I fully control for selection on beliefs in estimation using the rational expectations assumption and the selection structure of the model, which allows me to obtain consistent estimates without exclusion restrictions on cross-occupation experience.[10] 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.[11]

The rest of the article 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 higher returns than fostering entrepreneurship later in the life cycle. Section 8 concludes.

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

[8] 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, provided that entrepreneurial outcomes are more variable.

[9] I avoid assuming that individuals have certainty over a subset of their occupational ability.

[10] Since experience is endogenous to unobserved beliefs, setting the experience coefficient to 0 in an income equation assumes away selection concerns. I find strong evidence against this assumption.

[11] Under standard assumptions, the flow utilities without unobserved heterogeneity are exactly identified off long panels (Arcidiacono and Miller, forthcoming). Hence, it is unclear that one could separately identify unobserved heterogeneity in preferences (e.g., in risk aversion) from unobserved heterogeneity in ability/beliefs without making further structural assumptions or introducing exclusion restrictions.

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 PSID.[12] 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. 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.[13]

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).[14] 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 0 or negative. Hourly income is obtained by dividing annual income by annual working hours. Wealth data in the PSID are collected only on a few periods before 1996. Thus, I construct a measure of permanent wealth, denoted by ω, using all wealth observations available (including the ones after 1996), 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 of age. Median (mean) permanent wealth is 300 (400) thousands of dollars.[15]

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.[16] Furthermore, most of the 4,294 occupational spells in the sample are paid-employment spells, which are more than 60% 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 nine years of paid-employment experience on average. This opens a gap in average-entry age between entrepreneurial and salaried occupations that run opposite to the prediction of a parsimonious model of uncorrelated learning about ability

[12] I restrict the sample to men due to data availability in the early years of the survey, and to avoid modeling the intensive margin of labor participation and the relationship between entrepreneurship and fertility.

[13] 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.1.1). In the learning framework of this article, modeling missing yearly data as unobserved would imply cumbersome integration over long sequences of unobserved signals.

[14] In the sample, 79% (87%) of unincorporated (incorporated) individuals respond affirmatively to the business ownership question in the PSID, 41% (22%) 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 with more complex administrative activities and higher administrative costs.

[15] 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 life-cycle model of labor supply in MaCurdy (1981). The level of the permanent wealth measure depends on the degree of the age polynomial. The estimated value of the constant is about $418,000 (Appendix A.1.1).

[16] 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	Not Working
Total	4294	1707	1652	453	194	288
Percentage		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.

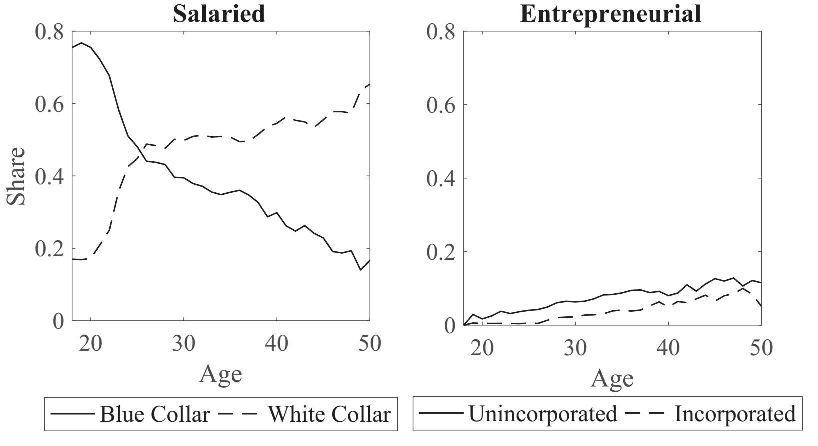


FIGURE 1

OCCUPATIONAL CHOICE: PARTICIPATION BY AGE

TABLE 3
TRANSITION MATRIX

	Blue Collar	White Collar	Unincorporated	Incorporated	Not Working
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
Not working	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 $t+1$.

(Miller, 1984). Notably, although Table 2 indicates that few individuals start their careers as entrepreneurs, Figure 1 shows that participation increases as individuals age.

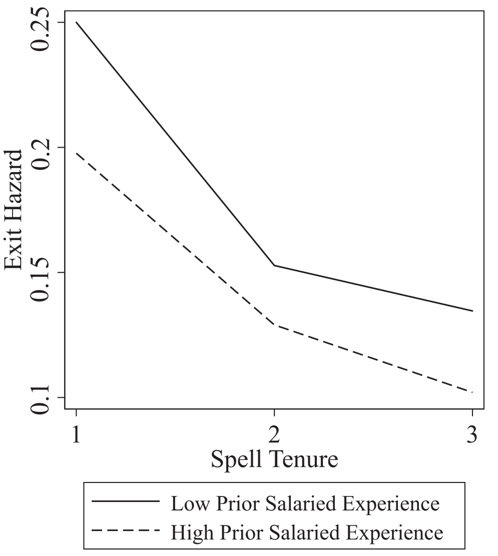
Different types of entrepreneurs vary in their patterns of transition into paid employment. Table 3 shows that unincorporated entrepreneurs are just as likely to switch to either blue collar or white collar work, whereas incorporated entrepreneurs tend to transition to white collar work. Consistent with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

TABLE 4

SAMPLE CHARACTERISTICS PER OCCUPATION

	All	Blue Collar	White Collar	Unincorporated	Incorporated	Not Working
Observations	21334	8902	9957	1403	602	470
Percentage	100.00	41.73	46.67	6.58	2.82	2.20
Age	31.04	28.92	32.21	33.93	36.94	30.45
	(7.27)	(6.65)	(7.18)	(7.35)	(7.00)	(8.06)
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
More than college	0.23	0.05	0.39	0.27	0.34	0.17
Hours worked	2147	2096	2234	2329	2703	
	(693)	(617)	(559)	(819)	(724)	
Hourly labor income	18.71	14.16	21.24	21.30	37.91	
	(13.72)	(7.24)	(12.29)	(21.04)	(47.19)	
Residual						
Hourly 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 ordinary least squares (OLS) regressions on race, education, and second degree polynomials of occupation-specific experience.



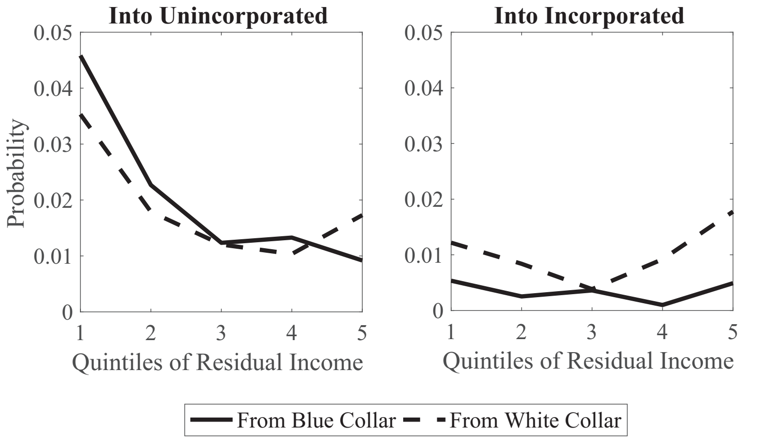
NOTES: Probability of switching out of first entrepreneurship spell in $t+1$ conditional on surviving until t, by prior salaried experience. Low (high) corresponds to <= (>) six years of salaried experience prior to first entrepreneurship spell.

FIGURE 2

EXIT HAZARD FROM FIRST SPELL OF ENTREPRENEURSHIP BY SALARIED EXPERIENCE

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.

Paid-employment experience is associated with entry, exit, and success in entrepreneurship. 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 5% significance that those with high



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

FIGURE 3

PROBABILITY OF SWITCHING INTO ENTREPRENEURIAL OCCUPATIONS

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.

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.[17] 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 to determine the degree of correlated learning.

In the remainder of the article, I build 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 model features and exemplify the sources of identifying variation in the data.

3. MODEL

In the model, forward-looking individuals face dynamic incentives that reflect accumulation of human capital 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 his acquired managerial skills into his business. Additionally, his success or failure as a manager may reveal his entrepreneurial ability. Individuals are risk averse and face income risk due to irreducible idiosyncratic uncertainty and uncertainty about their own ability, which they can reduce by observing their labor market outcomes. Their

[17] Residual income obtained from occupation-specific OLS regressions after controlling for demographics and a quadratic in occupation-specific experience.

productivity in each occupation is determined by their human capital, 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 while facing financial constraints that prevent them from fully insuring against income risk.

3.1. *Occupations, Income, and Individual Characteristics.* An individual in the model, denoted by 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 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. Occupation-specific experience, x_t, is a K-dimensional vector that partly determines his productivity in each occupation. He starts his career with no experience and acquires it over time as a function of his choices:[18]

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

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.[19] His hourly income in occupation $k > 0$ is not constrained to be positive (e.g., if an entrepreneur has losses), and is the sum of three components, a mapping $f_k(\cdot) : \mathbb{Z}^K_{\geq 0} \to \mathbb{R}$ of his experience vector in addition to his fixed, occupation-specific ability $\mu_k \in \mathcal{M}$, and a temporary productivity shock denoted by η_{kt+1}:

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

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

3.2. *Information Structure and Learning.* Both the distribution of occupational ability and the distribution 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 $\mathcal{I} = \varnothing$, then he only knows the population distribution of occupational ability.[20]

[18] There is evidence in the literature that human capital depreciates (e.g., Altuğ and Miller, 1998; Skira, 2015). Although the model in the main body of the article abstracts away from human capital depreciation, Appendix A.5.1 specifies and estimates an extended version with human capital depreciation. The main results remain largely unchanged.

[19] $\bar{L}_k$ is specified as the average number of hours worked by individuals in occupation k in the sample. Since men tend to work full time (Table 4), the model abstracts from the working hours decision given an occupation. Hence, preference for working in a given occupation will be net of the disutility from working.

[20] $\mathcal{I}$ 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).

The individual only observes a temporary, occupation-specific productivity shock after joining an occupation.[21] 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:

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

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.[22]

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 restrictions on $\mathcal{I}$ and $F_{\mathcal{M}}$, and does not impose which occupation is affected the most by information frictions. 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*.[23]

3.3. *Utility.* The individual is forward-looking and his discount factor for future utility is β. His flow utility $u(\cdot, \cdot; \rho) : \mathbb{R} \times \mathbb{R}^{K+1} \to \mathbb{R}$ depends on consumption c_t and a vector of observed, alternative-specific utility shocks ϵ_t drawn from a distribution characterized by the cumulative distribution function $F_\epsilon : \mathbb{R}^{K+1} \to [0, 1]$, which is known. His utility function is characterized by the risk aversion parameter ρ. Although the current piece is confined to homogeneity in risk aversion for tractability, it improves upon concurrent structural, life-cycle models of entrepreneurship that omit risk aversion altogether (Dillon and Stanton, 2017; Humphries, 2018).[24]

The utility function incorporates the fact that risk-averse individuals are less willing to try risky occupations, as entrepreneurial occupations are often thought to be. Risk aversion interacts with the information structure. On the one hand, the individual always dislikes believed future income variance caused by temporary productivity shocks. On the other hand, when his believed income variance is caused by occupational ability, he faces a trade-off between information value and risk aversion because, before knowing his place in the distribution, he likes occupations where there is high variance in ability. Moreover, if there is cross-occupation learning about ability, he may optimally choose to 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.

[21] 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.

[22] Given the definition of ζ_{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 A.2.1.

[23] Information spillovers are the reason why the problem cannot be represented in terms of dynamic allocation indices à la Miller (1984).

[24] Allowing for unobserved heterogeneity breaks the two-stage estimation process in Section 4 because the log likelihood (Appendix A.3.1) is no longer separable. In the literature, Gayle et al. (2015) do not find statistically significant heterogeneity based on observables among managers; in health insurance markets, Cohen and Einav (2007) find that unobserved heterogeneity is important but Handel (2013) does not find substantial unobserved heterogeneity.

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_ϵ 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 $N(\mathbf{0}, \Delta_n)$, allowed to vary by education level n since individuals with different education levels are likely selected based on ability. Occupation-specific productivity shocks are distributed $N(0, \sigma^2_{\eta_k})$. 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: $\mathbb{B}_{t_0} = \langle \mathbf{0}, \Delta_n \rangle$. Moreover, the normality of both $F_{\mathcal{M}}$ and F_{η_k} yields a posterior that 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 θ_k, is augmented to include all the components of h_t except age, 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 $T + 1$.[25] His flow utility $u(\cdot, \cdot; \rho)$ is characterized by a constant absolute risk aversion (CARA) function of consumption c_t where ρ is the absolute risk aversion.[26] The flow utility mapping is augmented to include some components of h_t, and the marginal contribution of consumption to his utility is occupation-specific and determined by the nonpecuniary cost of each occupation, α_{kt}. Concretely, the individual's flow utility from alternative k is:

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

where

$$\alpha_{kt}(h_t) = \exp\{\alpha_{k0} + \alpha_{k1}\mathit{black} + \alpha_{k2}\mathit{married}_t + 1\{x_{kt} = 0\}(\alpha_{k3} + \alpha_{k4}t + \alpha_{k5}\omega + \alpha_{k6}t\omega)\}. \tag{5}$$

Although for notational simplicity, I write $\alpha_{kt}(h_t)$, where α_{kt} includes a first-time entry cost that depends on permanent wealth (ω) and age. (Appendix A.4.5 shows that entry costs and nonpecuniary benefits can be treated as the indirect utility representation of cost terms in the budget constraint.) This is a reduced-form way of 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 through the permanent wealth measure. The interaction term between age and permanent wealth captures whether more affluent individuals are less affected by the savings mechanism proxied by age.

The setup for the consumption and wealth choices follows Margiotta and Miller (2000). Individuals have access to a contingent claims market for consumption goods that they use to

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

[26] Although the support of $u(\cdot, \cdot; \rho)$ admits negative consumption to guarantee an interior solution of the consumption smoothing problem, negative consumption is unlikely to occur since an individual's negative income (e.g., if his business losses money) is unlikely to be persistent over time, and he will smooth temporary negative shocks using the technology described below. Moreover, I do not model welfare or unemployment programs that would guarantee a consumption floor. Finally, individuals in reality may become less risk averse as they amass wealth. I abstract from this mechanism to avoid further complexity using a CARA specification that yields substantial gains in tractability (Proposition 1).

smooth consumption. Let λ_τ denote the derivative of the price measure for claims to consumption at date τ.[27] 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 ξ_{t+1} and reflects the allocation of financial resources between current-period consumption and next-period savings.

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

Income is assumed uninsurable, capturing unobservable insurance risk and unobserved levels of effort in labor supply (Green, 1987). For instance, once contingent claims have been traded, individuals may have no incentive to actually work in the occupation they selected or to exercise effort in their enterprises. Public disclosure of the sources of occupational income ensures that contracts are written so as to nullify such incentives, preventing individuals from being able to fully insure and receive their expected income with certainty. This financial constraint affects primarily participation in entrepreneurial occupations that offer higher income risk. Given the curvature of the utility function, the constraint binds more strongly for individuals with high expected ability but who remain fairly uncertain about it (high belief variance) and who face high entry costs that they cannot fully finance.

The model differs from a standard framework of entrepreneurship entry and liquidity constraints an lá Evans and Jovanovic (1989) in that it does not include the role of financial capital in the production function of entrepreneurial income, and it does not specify a financial capital constraint. In the standard model, constrained entrepreneurs who have high ability and low wealth attain a suboptimal level of capital investment in their businesses, which decreases their net income. In the framework, here entrepreneurial income as a function of ability is unaffected by financial capital upon entry, and individuals pay a cost to enter into entrepreneurship that depends on their education, age, and permanent wealth.

Another difference relative to the framework in Evans and Jovanovic (1989) is that individuals in my model are forward-looking in their use of financial resources and smooth consumption over time. This is another important feature that sets the framework apart from recent structural models of entrepreneurship that impose hand-to-mouth consumption (Dillon and Stanton, 2017; Humphries, 2018). Instead, my model captures consumption choices while preventing the econometrician from understating the effects of risk aversion. For instance, since entrepreneurship is a 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.

3.5. *Optimal Choices.* Upon retirement, the individual simply smooths his remaining wealth; the solution of this part of the problem is analytical and is presented in Margiotta and Miller (2000). At the beginning of any period before retirement, the individual receives his income from last period's occupation and observes his vector of taste shifters ε_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. As is common in the literature that uses contingent-claims to represent an economy with uncertainty (e.g., Altuğ and Labadie (1994)), given the CARA nature of the flow utility, the consumption smoothing problem satisfies a portfolio separation property according to which the individual's optimal consumption and savings choices can be written in terms of a bond b_τ and a security a_τ.[28] Individuals in the model accurately forecast the price of both assets.

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

[28] Concretely, b_τ and a_τ are given by:

$$b_\tau \equiv E_\tau\Big[\sum_{s=\tau}^{\infty} \tfrac{\lambda_s}{\lambda_\tau}\Big] \qquad a_\tau \equiv E_\tau\Big[\sum_{s=\tau}^{\infty} \tfrac{\lambda_s}{\lambda_\tau}(\ln \lambda_s - s\ln\beta)\Big],$$

Proposition 1 below shows that the value function of the problem can be separated into two factors: an indirect utility function for wealth and an index that captures the value of human capital and information. It provides the ex ante value function of an individual—that is, the value function before knowing the realization of the vector of taste shocks, ε_t—at any age before retirement, when he is choosing an occupation in addition to his consumption and asset portfolio. Let $\tau(t)$ be the calendar date when the individual is of age t.

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

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

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

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

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

The probability of choosing k at age t conditional on characteristics and beliefs is denoted by $p_{kt}(h_t, \mathbb{B}_t)$. ε^*_{kt} is the value of the taste shock ε_{kt} conditional on option k being chosen at t. The deterministic transition from h_t into h_{t+1} is denoted by $\bar{H}_{kt+1}(h_t)$, 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 A.2.2. ■

The index $A_t(h_t, \mathbb{B}_t)$ in Proposition 1 is a strictly positive average of expected outcomes weighted by the ccp 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 corresponds to the permanent increase in consumption that can be attained by using annual income to purchase bonds. It scales the future utility from human capital and information capturing the effect of consumption smoothing concerns on occupational choices. Higher values of the prior mean or higher values of human capital are associated with lower values of the index $A_t(h_t, \mathbb{B}_t)$ 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 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 nonpecuniary costs, α_{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

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

where b_τ is a bond that, contingent on the history through calendar date τ, pays a unit of consumption from period τ in perpetuity in date-τ prices. a_τ is a security that pays the random quantity $(\ln\lambda_s - s\ln\beta)$ of consumption units from period τ in perpetuity, in date-τ prices. 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.

His occupational choice accounts for consumption smoothing but 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.[29] Trade-offs between occupations are characterized by differences in idiosyncratic utility shocks, ε_{kt}, differences in nonpecuniary benefits and entry costs, $\alpha_{kt}(h_t)$, 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, $A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})$. 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 between occupations 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*

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

where

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

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 $d \in \{d_{kt}, d_{0t+1}, d_{0t+2}, \ldots, d_{0T}\}$.

PROOF. See Appendix A.2.3. ■

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). Higher nonpecuniary and entry costs and lower expected utility from compensation, potentially 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 θ_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, because beliefs are unobserved, individual-specific, and time-varying (Gibbons et al., 2005). My strategy consists of using the panel dimension of the data to model the selection process: I specify how beliefs are formed and the mechanism through which they affect occupational choices. However, similar paths

[29] This separability property breaks in the standard credit constraints model where wealth levels can prevent individuals from obtaining the optimal business scale, rendering their occupational choice dependent on current wealth.

of choices may be driven by multiple combinations of heterogeneous initial beliefs. To attain identification, I assume that individuals have rational expectations (James, 2011) and that at the beginning of their careers, they have received no individual-specific information about ability that is not captured by their education level. These assumptions anchor initial beliefs so that all individuals of a given level of education start off their careers with beliefs that are equal to the distribution of ability in their respective populations. Then, I can use panel data on occupation and income and the Bayesian learning structure to control for selection on beliefs.

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 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. Notably, since education indicators enter linearly in the income equations, the mean of the distribution of ability, which is education-specific, is not identified and is normalized to 0.[30] Finally, the variance of the distribution of productivity shocks is identified from the excess variation in residual income per occupation.[31]

Identification of the flow payoffs follows from results in Magnac and Thesmar (2002) and Arcidiacono and Miller (forthcoming). 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 0 in each state and time period—in other words, $\alpha_{0t}(\cdot) \equiv 1$. Hence, the functional form assumptions regarding the utility function provide overidentifying restrictions; for instance, since initial priors are anchored by the rational expectations assumption, there are overidentifying assumptions for the risk aversion parameter.[32]

Estimation of the structural model entails a combination of an EM algorithm and a ccps estimator (Hotz and Miller, 1993; Arcidiacono and Miller, 2011; James, 2011). Because individuals select on beliefs instead of 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.[33] The likelihood of the data and a more detailed description of the estimation procedure are presented in Appendix A.3. Standard errors are corrected for the two-stage estimation procedure using subsampling estimation over 100 subsamples without replacement.

First stage. An EM algorithm yields estimates of the collection of income parameters, denoted as Θ, and the covariance matrix of the population ability distribution conditional on education level n, denoted as Δ_n.[34] 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, $\mathcal{M}$. Fast estimation of income and ability parameters is obtained by avoiding multidimensional integration over unobserved ability vectors.

Second stage. A ccp estimator yields estimates of the collection of parameters of the utility function, denoted by Υ. The estimator uses Proposition 2 that provides a mapping from future choice probabilities and utility parameters into current choice probabilities. An iterative

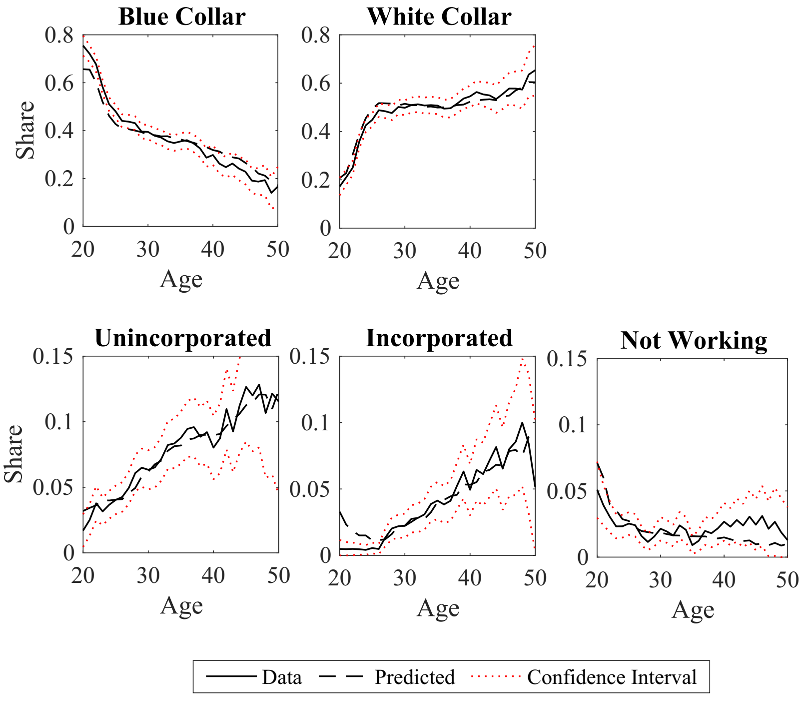
[30] Alternatively, the education-specific shifters in the income equation can be thought of as education-specific means for the ability distribution.

[31] This can be seen more clearly in the updating rules of the estimation algorithm in Appendix A.3.2.

[32] In fact, identification of the risk aversion parameter could rely on the panel dimension of the data. This is because 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 in choices would help identify the risk aversion parameter.

[33] Additive separability requires that no measurement error is assumed in the hourly income data. Integration over measurement errors breaks additive separability.

[34] Θ includes the variance parameters of the productivity shocks, $\sigma^2_{\eta k}$.



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

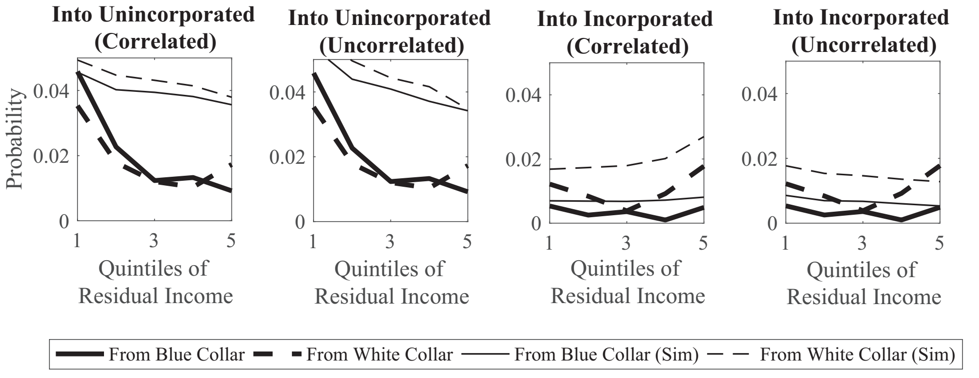
FIGURE 4

MODEL FIT [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]

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).[35] 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 22. Additional fit statistics in Appendix A.4.3 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 it is not targeted in estimation, the model 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.

[35] The asymptotic properties of iterative estimators where the first-iteration estimator is consistent (as is the case here) are discussed in Amemiya (1985).



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.

FIGURE 5

PROBABILITY OF SWITCHING INTO ENTREPRENEURIAL OCCUPATIONS

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 6(a) illustrates estimated returns to own-occupation experience.[36] It shows that the returns to blue collar experience are rather flat, whereas the returns to incorporated experience are the steepest. Figure 6(a) suggests that individuals aspiring to become highly productive incorporated entrepreneurs must accumulate at least five years of incorporated entrepreneurial experience. Additionally, learning-by-doing can happen across occupations. Figure 6(b) illustrates estimated returns to cross-occupation experience. It shows that although expertise in entrepreneurship, as measured by high accumulated experience, increases productivity in paid employment, low levels of entrepreneurial experience can reduce it.[37]

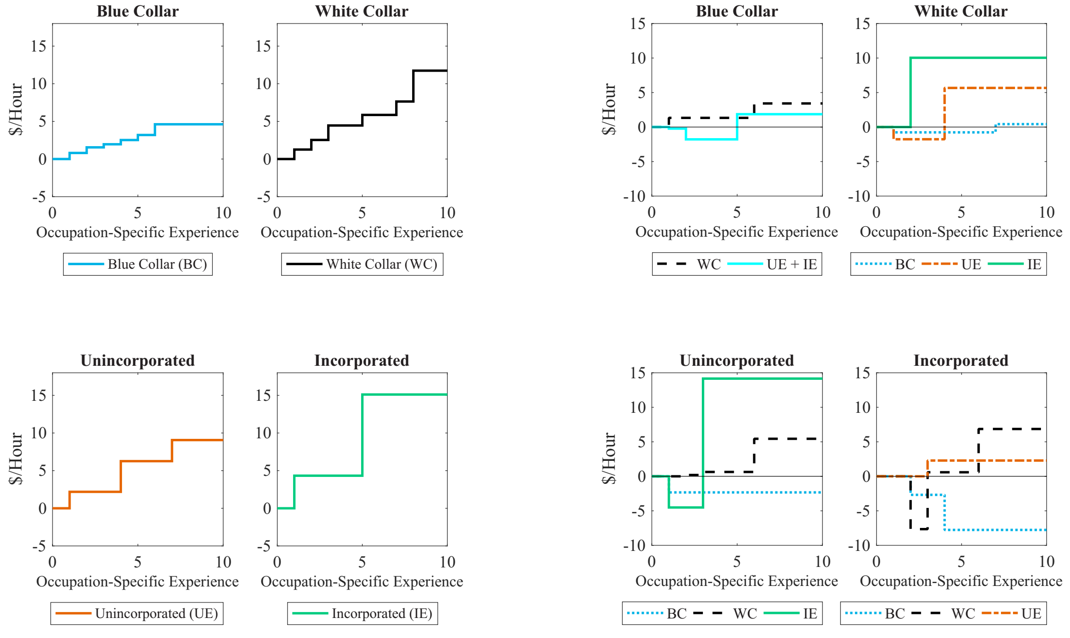
As they work, individuals also learn about their occupation-specific abilities. This learning process is determined by the covariance matrix of their ability distribution. Results suggest that there is more for individuals to learn about in entrepreneurship. In other words, there is higher variation in entrepreneurial ability than in paid-employment ability. Consider the covariance matrix for individuals with more than college education in Table 5.[38] Since one unit of ability is equivalent to \$10 in hourly wages, one standard deviation above the population mean translates into $10 \times \sqrt{10.88} \approx \$33$/hour for incorporated ability and into $10 \times \sqrt{0.87} \approx \$9$/hour for white collar ability. In addition, since more education tends to be associated with higher variation in ability (Table A.3), 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 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.

There is an incentive for young individuals at the beginning of their careers to attempt entrepreneurship to learn whether they are high ability. Alternatively, individuals can use

[36] Estimates for Equation (2) are presented in Table A.2.

[37] 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.

[38] This result is seen across education levels (Table A.3).



(a) Own-Occupation Returns (b) Cross-Occupation Returns

NOTES: Returns implied by estimates of Equation (2) that is specified using step functions (Table A.2). Occupations in Figure 6(b) are: white collar (WC), blue collar (BC), unincorporated entrepreneurship (UE), and incorporated entrepreneurship (IE). In blue collar work, experience from both entrepreneurial occupations is pooled. 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.

FIGURE 6

RETURNS TO EXPERIENCE [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]

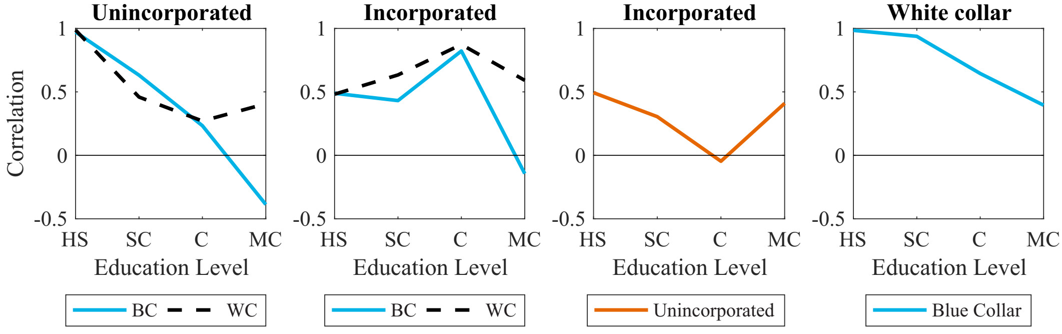
TABLE 5

POPULATION ABILITY COVARIANCE MATRIX (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: This table includes point estimates (coeff) and standard errors (se) corrected for two-stage estimation using subsampling estimation over 100 subsamples without replacement. Covariance matrix of the joint distribution of unobserved ability conditional on education, denoted by Δ_n.

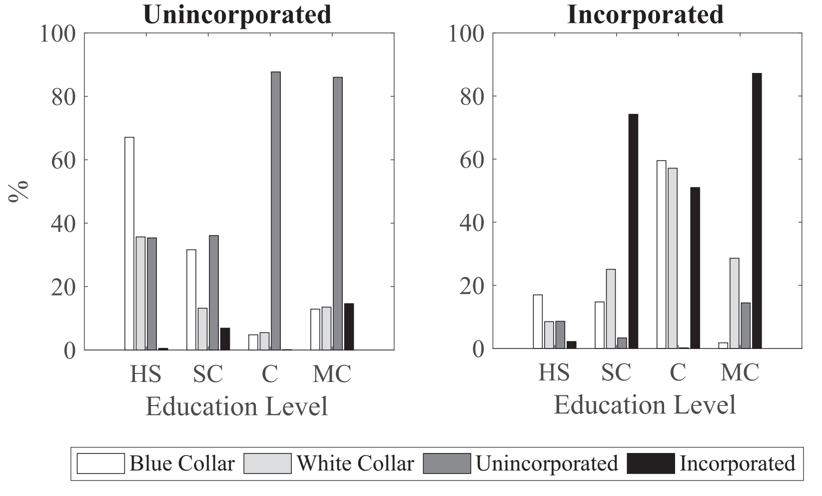
their white collar success as an indicator of entrepreneurial ability. However, the informational value of entrepreneurship is undermined by higher idiosyncratic variation in entrepreneurial outcomes (Table A.4), which can slow down learning. As an assessment of how fast own- and cross-occupation learning about ability can happen, Figure 8 displays the percentage of prior uncertainty about entrepreneurial ability that is eliminated after working for five years in each occupation. For individuals with more than college education, uncertainty about incorporated ability decreases by almost 90% after five years of incorporated experience and decreases by about 30% after five years of white collar experience. Surprisingly, for the college educated, paid



NOTES: Correlation between abilities per education level implied by estimates in Table A.3. 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, respectively. The third figure shows the correlation between entrepreneurial abilities. The fourth figure shows the correlation between salaried abilities.

FIGURE 7

IMPLIED CORRELATION BETWEEN ABILITIES [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]



NOTES: Percentage of prior variance about ability in occupation $k \in$ {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 A.3 and A.4.

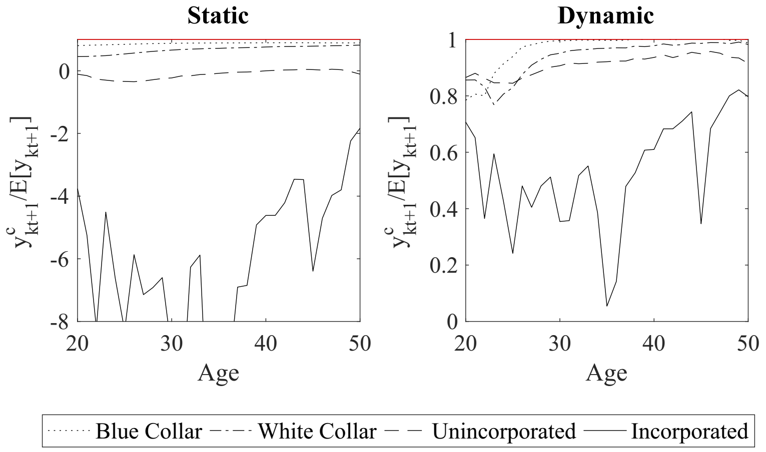
FIGURE 8

PRIOR VARIANCE ELIMINATED AFTER FIVE YEARS OF OCCUPATION-SPECIFIC EXPERIENCE

employment is a slightly better source for learning about incorporated ability than incorporated entrepreneurship itself. The same result does not hold for unincorporated entrepreneurship.

5.2. *Risk Aversion, Costs, and Benefits.* This section discusses the estimates of risk aversion, entry costs, and nonpecuniary benefits. In presenting results, the list of point estimates is deferred to Table A.5 and utility parameters are discussed using a certainty equivalent exercise and monetary equivalents for entry costs and nonpecuniary benefits.[39]

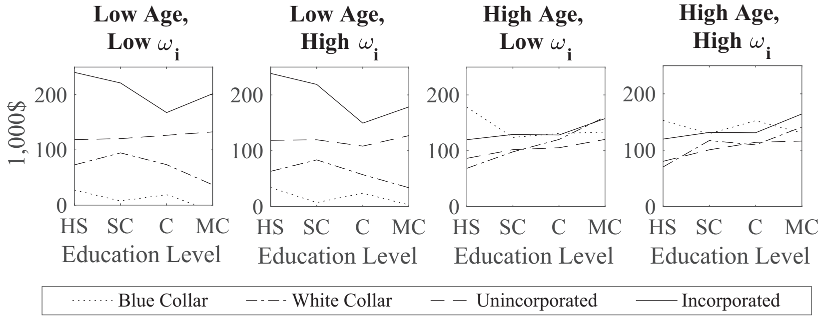
[39] Risk aversion and most entry cost and nonpecuniary benefit parameters are statistically significant. Appendix A.4.1 shows that the point estimate of absolute risk aversion is within the spectrum of results found in the previous literature. Individuals in my sample are more risk averse than those in Gayle and Miller (2009) and Gayle et al. (2015) and less risk averse than those in Barseghyan et al. (2013) and Handel (2013).



NOTES: On the y-axis is the average of the ratio of the certainty equivalent y^c_{kt+1} (Appendix A.4.4) 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.

FIGURE 9

CERTAINTY EQUIVALENT [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]



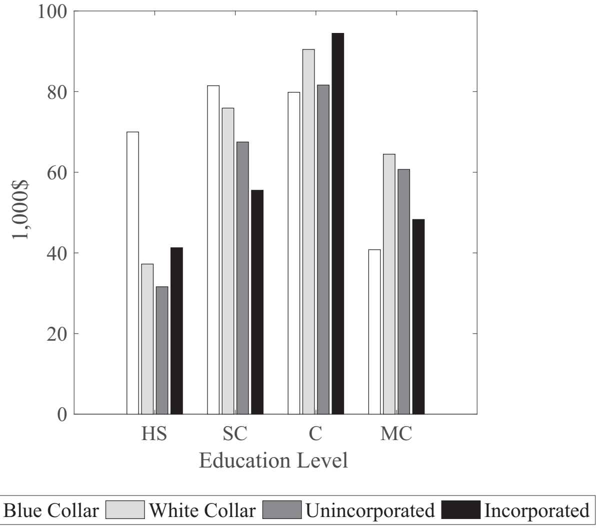
NOTES: On the x-axis of each panel is the education level: high school (HS), some college (SC), college (C), and more than college (MC). On the y-axis is the monetary equivalent of entry costs, obtained using the estimates in Table A.5 and Equation (A.59) both in Appendix A.4. Ages are 20 and 40. Permanent wealth ω_i points are the 10th and 90th percentiles.

FIGURE 10

MONETARY EQUIVALENT OF ENTRY COSTS

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 entrepreneurs to value expected benefits associated with future human capital and future information. This dynamic consideration mitigates the effects of risk aversion.

In addition to risk aversion, first-time entry costs constitute a barrier to entrepreneurship, in particular to young entrepreneurship. Entry costs decrease with age as well as with permanent wealth. Although the permanent wealth coefficient captures lifetime affluence, the age coefficient aims to capture the wealth accumulation mechanism as younger individuals will have had less time to accumulate resources and will have weaker credit histories. The interaction between age and permanent wealth is positive, which suggests that more affluent individuals are less affected by the savings mechanism proxied by age. Figure 10 introduces monetary equivalents of



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 nonpecuniary benefits not related to entry, obtained for a white, married man using the estimates in Table A.5 and a similar derivation as in Equation (A.59).

FIGURE 11

MONETARY EQUIVALENT OF NONPECUNIARY BENEFITS

entry cost estimates. On the one hand, the monetary equivalent of entry cost into incorporated entrepreneurship is about \$50,000 higher for individuals age 20 than for individuals age 40.[40] On the other hand, 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.[41] Since I estimate incorporated ability to vary much more widely than ability in other occupations, individuals with high expected incorporated entrepreneurial ability and high uncertainty around it—the latter being more prevalent among the young—are more financially constrained as they cannot fully insure income risk; this result is consistent with Levine and Rubinstein (2018). Overall, my results are consistent with the credit constraints mechanism.

Nonpecuniary motivations can also affect the decision to become an entrepreneur (Hamilton, 2000; Hurst and Pugsley, 2017). Monetary equivalents of nonpecuniary benefits in Figure 11 suggest that entrepreneurial nonpecuniary benefits do not always dominate salaried nonpecuniary benefits once dynamic considerations are introduced.[42] 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 nonpecuniary benefits of incorporated entrepreneurship for the college educated are higher than those from any other occupation at any education level.

Although the framework in this article captures many economic determinants of the choice to be an entrepreneur versus a paid employee, it also has limitations. Motivated by the lack of wealth data in my sample period, modeling assumptions yield an analytical solution to the individual's savings decision that renders the occupational choice independent of current wealth. In other words, savings choices affect occupational choices only through optimal consumption smoothing (Propositions 1 and 2). However, since individuals in reality may not be able to

[40] 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.

[41] A counterfactual exercise (Table A.12) 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.

[42] These results are in line with those in Dillon and Stanton (2017).

smooth their consumption as much, in Appendix A.5.2, I use the only three years of wealth data available during my sample period (1984, 1989, and 1994) to explore in reduced form the relationship between current wealth and the decision to switch into entrepreneurship, comparing the role of current wealth with that of several variables included in the model. The descriptive results in Appendix A.5.2 suggest that current wealth increases (decreases) the probability of switching into incorporated (unincorporated) entrepreneurship. The marginal effect from current wealth on the probability of switching is lower than the effect from one year of previous entrepreneurial experience and higher than the effect from a one standard deviation in beliefs about entrepreneurial ability. The model in Section 3 attempts to partially capture the role of current wealth by interacting occupational entry costs with age, as younger individuals are likely to face higher barriers to enter into entrepreneurship due to their lower wealth on average (Figure A.1). However, the age profile of entry costs could be contaminated if younger individuals also differ in their risk aversion.

Employer-sponsored health insurance (ESHI), which could provide incentives for individuals not to venture into entrepreneurship (Holtz-Eakin et al., 1996; Fairlie et al., 2011), is not included in the framework. Although this effect may be less pronounced for young individuals (Hahn and Yang, 2016), omitting this mechanism is likely to bias downward the estimates of nonpecuniary benefits for entrepreneurship.[43] In Appendix A.5.3, I use the only year in the sample period that contains information on ESHI (1984) to explore its relationship to the decision to become an entrepreneur. The results suggest that ESHI decreases the probability of switching into entrepreneurship, especially if currently engaged in blue collar work. However, the marginal effect of ESHI on the probability of switching does not dominate the marginal effect of variables associated with learning by doing (prior experience) and learning about ability (beliefs).

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.[44] Results below include both types of entrepreneurs, but the discussion focuses on incorporated entrepreneurs—recent literature suggests that entrepreneurship is better proxied by the incorporated, whereas self-employment is better proxied by the unincorporated (Levine and Rubinstein, 2017, 2018). 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 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.[45]

[43] Morrisey et al. (1994) find that incorporated businesses are more than twice as likely as unincorporated businesses to provide health insurance coverage. Hence, this bias is likely to affect more the unincorporated.

[44] The comparison requires solving the model and simulating data under the baseline and each of the counterfactuals, keeping the initial state fixed. Appendix A.4.2 shows how the model is solved backward using the representation in Proposition 2. Appendix A.4.6 further explains the counterfactual regimes including extended results in Table A.12.

[45] Let $R_k(x)$ be the return to experience in occupation k for somebody who has worked x years in occupation k and zero years in any other occupation (Figure 6a). The fixed hourly return to observed skill for individuals in occupation

- $\hookrightarrow$ *No learning about ability (full information)*: Individuals know their ability vector $\mathcal{M}_i$ but the initial level of uncertainty (i.e., income risk) remains unchanged.[46]
- $\hookrightarrow$ *No cross-occupation learning-by-doing*: Productivity in one occupation is invariant to experience in another.
- $\hookrightarrow$ *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 Δ_n diagonalized.
- $\hookrightarrow$ *No uncertainty*: There is no income risk; individuals know their ability vector $\mathcal{M}_i$ and there is no idiosyncratic variation in hourly income.
- $\hookrightarrow$ *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.

Several forces prevent individuals from attempting entrepreneurship: having to learn how to be one, uncertainty in ability, risk aversion, and entry costs. Figure 12(a) 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.[47] The effect of income risk is measured as the difference in ratios from providing full information versus eliminating all income risk, which amounts to about 40% for incorporated entrepreneurship.[48] Flattening the entry cost profile has a similar effect to eliminating all uncertainty: Individuals would be 75% more likely to attempt incorporated entrepreneurship.[49]

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 either their entrepreneurial productivity or their entrepreneurial beliefs using their paid-employment experience and outcomes. However, the role of cross-occupation learning may disappear once human capital depreciation, which is not included in the model, is accounted for. In a model with human capital depreciation, all cross-occupation experience loses value overtime after an occupation switch. Moreover, human capital depreciation can decrease the incentives to change occupations after fine-tuning beliefs, potentially undermining the role of cross-occupation learning about ability as well. Appendix A.5.1 extends the baseline model to allow for human capital depreciation, estimates the extended model, and undertakes two counterfactual decompositions under the extended model: one that shuts down cross-occupation returns to experience and one that shuts down cross-occupation learning about ability. The results indicate that the parameters characterizing the learning about ability structure and the utility parameters remain largely unchanged. The counterfactuals with human capital depreciation reveal that the role of cross-occupation learning

k under this counterfactual regime is:

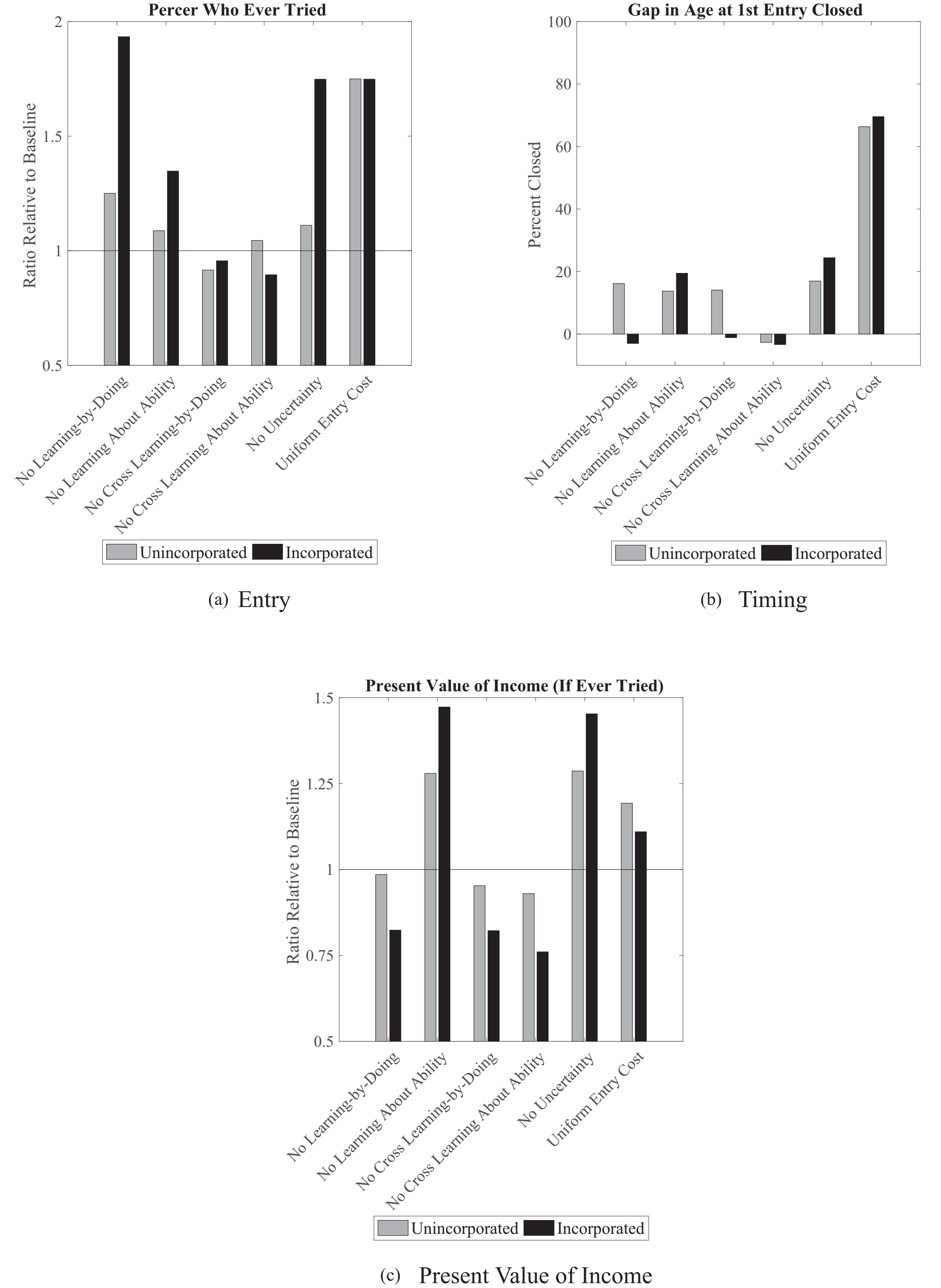
$$\bar{y}_k = \frac{1}{20}\sum_{x=1}^{20} R_k(x).$$

[46] In terms of Equation (2), this amounts to changing the value of the idiosyncratic income variance in occupation k from just σ_{η_k} to $\sigma_{\eta_k} + \Delta_{n,\{k,k\}}$, where $\Delta_{n,\{k,k\}}$ is the kth term in the diagonal of Δ_n.

[47] Results in Appendix A.4.6 show that under full information, participation and spell duration decrease for salaried occupations as they increase for entrepreneurial occupations. For example, under full information, the share of individuals who attempt the blue (white) collar occupation during their careers decreases by 6% (3%).

[48] Recall that in the full information, counterfactual individuals know their ability vector but their initial level of uncertainty remains unchanged. This is how the effect of income risk is separated from the effect of sorting on ability.

[49] Given the results in Table A.5, flattening the entry costs with respect to age amounts to young individuals facing a lower entry cost than in the baseline. For comparison, Table A.12 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.



(a) Entry

(b) Timing

(c) Present Value of Income

NOTES: Comparison of simulated data from the baseline model versus several counterfactual regimes. Gap in first-entry age measured relative to average first-entry age into white collar.

FIGURE 12

by doing and cross-occupation learning about ability, and hence the value of this type of learning, is not eliminated in the presence of human capital depreciation.

Young entrepreneurs are harder to find because young individuals face higher entry costs and have less information about their entrepreneurial ability. The gap in first-entry ages between entrepreneurial occupations and salaried occupations is about 10 years (Table 1). Figure 12(b) shows the percentage of the gap, relative to white collar, which 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%. Eliminating all income risk has a smaller effect on the gap in first-entry ages, as 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 12(c) 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 PVI that is about 50% higher. Flattening entry costs also increases the PVI of the incorporated, although only by about 10%, because successful young entrepreneurs who 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, I study counterfactual policies fostering young incorporated entrepreneurship that target these barriers. Subsidies and entrepreneurship education are both effective in increasing young entrepreneurship and offer positive returns.[50] Using back-of-the-envelope calculations to bound education costs, I show that entrepreneurship education programs can offer higher returns than subsidies. Finally, I simulate the introduction of policies at older ages and find evidence, suggesting that the focus of policymakers on young entrepreneurs is not misplaced.

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 capturing long-term effects. 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.

On the long run, the subsidies increase the share of individuals who attempt incorporated entrepreneurship in their careers between 2% and 7% points and increase average gross PVI over all incorporated entrepreneurs and over all individuals. However, this may simply be reflecting the transfer of funds to young entrepreneurs. To better gauge the economic returns

[50] Appendix A.4.8 discusses how the differences between my framework and a standard framework of entrepreneurship entry and liquidity constraints a lá Evans and Jovanovic (1989) may affect these counterfactual policies.

TABLE 6
POLICIES FOSTERING YOUNG INCORPORATED ENTREPRENEURSHIP

		Subsidy ($1000s)		Entrepreneurship Education (Noise Variance Scale κ)			
	Baseline	25	50	10	5	2	1
Young entrepreneurs							
Tried in first five years	0.02	0.03	0.05	0.12	0.11	0.10	0.08
Mean belief ($ per hour) at first entry	3.7	2.5	1.2	82.4	75.1	61.3	48.4
Mean ability ($ per hour) at first entry	5.0	3.6	2.1	1.3	2.1	5.5	10.7
Bias (beliefabiliity)	−1.3	−1.1	−0.9	81.1	73.0	55.8	37.7
All entrepreneurs							
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
Gross PVI ($1000s)	757	773	780	944	941	961	983
Mean belief ($ per hour) at first entry	6.4	6.2	5.2	45.9	41.8	33.8	26.5
Mean ability ($ per hour) at first entry	5.5	5.4	4.6	2.0	2.3	4.0	6.6
Bias (beliefabiliity)	0.9	0.8	0.6	43.9	39.5	29.8	19.9
All individuals							
Gross PVI ($1000s)	508	513	526	573	575	578	581
Cost per capita ($1000s)		0.5	2.1	–	–	–	–
Policy returns							
Return $((PVI - PVI_{baseline})/Cost)$		11.2	8.4	–	–	–	–
Max cost for return as high as S25 ($1000s)		–	–	5.8	5.9	6.3	6.5
Max cost for return as high as S50 ($1000s)		–	–	7.7	7.9	8.4	8.7

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. The *Policy Returns* rows include per capita cost upper bounds for education policies to generate returns as high as those from the $25,000 and $50,000 subsidies (S25 and S50, respectively). *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, σ_υ, expressed as a scaled version of the estimated noise variance of trying incorporated entrepreneurship: $\sigma_\upsilon^2 = \kappa \cdot \sigma_{\eta 4}^2$.

of the subsidies, I compute the policy return as the gain in average PVI (computed over all individuals) under the policy relative to the baseline and divide it by the per capita cost of the policy. Returns from the subsidies are between 8.4 and 11.2 dollars per dollar invested.

Computing subsidy returns over the life cycle reveals that focusing on average ability at first entry, which decreases with the subsidies, may be misleading as it would suggest that higher participation is simply achieved at the expense of lower entrepreneurial quality (Shane, 2009; Hamilton et al., 2019). Instead, the dynamic model suggests that policies that relax entry costs and attract marginal entrepreneurs offer high returns. This result is driven by the wide dispersion in the distribution of entrepreneurial ability. Young individuals who decide to experiment with entrepreneurship as a consequence of the relaxation of the entry barrier generate on average more income over their life cycle than they would have if they had not experimented.

7.2. *Entrepreneurship Education.* Many policies that attempt to foster young entrepreneurship focus on entrepreneurship education.[51] 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

[51] 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. The OECD recommends that training approaches to foster young entrepreneurship should include experimentation (OECD, 2013).

individuals' elicited beliefs and intentions (Souitaris et al., 2007; Oosterbeek et al., 2010; von Graevenitz et al., 2010). However, in the words of von Graevenitz et al. (2010), "we have no means to assess how costly the mistakes of choosing the 'wrong' career would be to the students and to society at large (and hence) we cannot quantify the true economic and societal impact of entrepreneurship training." 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. Using my dynamic framework, I assess the value of entrepreneurship education policies that provide information signals.[52] Since my data do not include information regarding the signal quality of actual education programs or regarding program costs, I bound these results to make them comparable with subsidy outcomes. First, I use a mapping between my framework and findings in the literature to calibrate the position of an entrepreneurship program in the information quality spectrum. Second, I provide upper bounds for per capita program costs that guarantee policy returns as high as those generated by the subsidies. Finally, I compare these bounds to college and MBA course costs.

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 increases initial heterogeneity in entrepreneurial beliefs that will depend on ability as well as on luck. Notably, because abilities are correlated, this policy increases heterogeneity in beliefs across all occupations. Formally, the entrepreneurship education program yields for every individual a signal ζ_{4i}^{p} about his incorporated ability ($\mu_{4,i} \in \mathcal{M}_i$) given by

$$\zeta_{4i}^{p} = \mu_{4,i} + \nu_i, \qquad \nu_i \sim \ i.i.d.\ N(0, \sigma_\nu^2). \tag{12}$$

Individuals use the information contained in ζ_{4i}^{p} to update their beliefs before entering the labor market. It is assumed that no entrepreneurship education program can provide better information than becoming an entrepreneur for one period in reality. Hence, the noise variance from this intervention is bounded below by the idiosyncratic variance $\sigma_{\eta_4}^2$ (Table A.4) and can be written as

$$\sigma_\nu^2 = \kappa \cdot \sigma_{\eta_4}^2, \qquad \text{with } \kappa \geq 1. \tag{13}$$

The *Entrepreneurship Education* columns of Table 6 present the effects of policies that differ in their information quality ($\kappa \in \{1, 2, 5, 10\}$). The lower the quality (higher κ), the higher the percentage of young entrepreneurs and the bias in their beliefs at entry. Table 6 also shows a decline in average ability of young entrepreneurs for information quality below 50% ($\kappa \geq 2$). 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 a negative bias of \$1.3/hour in the baseline to a positive bias of \$81/hour from entrepreneurship education that provides 10% information quality ($\kappa = 10$). Although these results relate to the 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 different analysis of the information received (March and Shapira, 1987). Instead, overestimation at first entry emerges endogenously from uninformed, rational individuals who receive large positive signals due to luck.

[52] In this article, I have assumed homogeneity in risk aversion. However, if there is heterogeneity in risk aversion, individuals may react differently to a policy that provides information. For instance, the policy may attract individuals who are more risk averse on the margin. In Appendix A.5.4, I impose heterogeneity in risk aversion around the point estimate of ρ and find that although this "risk selection" mechanism is active, the results from the policy remain largely unchanged.

Entrepreneurship education provides long-term gains even when the quality of information is low ($\kappa = 10$). 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%. Notably, entrepreneurship education benefits all individuals, not only those who eventually become entrepreneurs. Relative to the baseline, average PVI increases between \$65,000 ($\kappa = 10$) and \$73,000 ($\kappa = 1$).[53] Hence, provided that the per capita costs of the policy are sufficiently low relative to the increase in average PVI—which back-of-the-envelope calculations below suggests might be the case—a policymaker who funds the program with a nondistortionary tax is willing to pay at most \$8,000 per capita to move from a program that generates low-quality information ($\kappa = 10$) to a program that generates high-quality information ($\kappa = 1$).

Next, I use findings in the literature to calibrate the information quality of an actual 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 into changes in the standard deviation of aggregate beliefs, and calibrate the value of κ to match the change in aggregate beliefs found in their paper (Appendix A.4.7). According to this mapping, the program's information quality is around half the quality of the information in the market ($\kappa = 2.1$), which implies that the program, if implemented universally, would increase young incorporated entrepreneurship from 2% to approximately 10%, raise the percentage of people who experiment with incorporated entrepreneurship in their careers from 15% to about 25%, and add around \$70,000 to the PVI of the average individual (Table 6, column $\kappa = 2$).

Finally, to compare these results against those from subsidizing entrepreneurship, the last two rows of Table 6 present the maximum per capita program cost, at every information quality level, that would make the returns from entrepreneurship education policies as large as those from subsidies. Upper bounds for per capita program costs vary between \$5,800 and \$8,700. Using data from the National Center for Education Statistics, a back-of-the-envelope calculation indicates that the average cost (tuition and required fees) in 2018 for an undergraduate, two-semester, six credits course at a four-year institution (public or private) is around \$2,800, less than half the minimum upper bound; a similar course at the graduate level in a private institution has an average cost of around \$4,500. Additionally, using data from *U.S. News*, I find that in 2019, a similar course in a top-20 MBA has an average cost of \$11,700, and \$7,500 in a school below the top 20 but above the top 50.[54] Hence, it seems plausible that entrepreneurship education programs aimed to foster young entrepreneurship offer higher returns than subsidies, although less so when implemented at more expensive, elite MBA programs.

7.3. *Why Young Entrepreneurs?* Although recognized international institutions recommend fostering young entrepreneurship (OECD, 2013), it is unclear what the returns of this approach are relative to fostering entrepreneurship at other stages in a person's career.[55] On the one hand, individuals who are uncertain about their entrepreneurial ability can benefit from education that provides early information because it allows them to finesse their participation decision and enjoy the benefits of better selection for longer (Miller, 1984) and because the marginal value of information is higher when individuals are fairly uninformed. On the other, due to correlated

[53] The gain in expected PVI from entrepreneurship education of 10% of quality relative to the baseline goes from \$700 for high schoolers to about \$200,000 for individuals with more than college education, which reflects the differences in entrepreneurial potential by education level.

[54] Costs in constant dollars of 2000. Cost per credit was obtained by assuming that tuition and required fees cover two semesters of full-time enrollment (12 credits per semester). In-state and out-of-state tuition costs were averaged. For undergraduate and graduate average tuition and fees data, visit: https://nces.ed.gov/programs/digest. For top 50 MBA tuition data, visit: https://www.usnews.com/best-graduate-schools/top-business-schools/mba-rankings.

[55] An alternative motivation for fostering young entrepreneurship is that it helps individuals avoid unemployment if entry into salaried employment is more difficult for inexperienced individuals.

TABLE 7
RATIO OF RETURNS FROM FOSTERING OLD VERSUS YOUNG ENTREPRENEURSHIP

Policy	Age 30	Age 40
Subsidy	0.86	0.30
Entrepreneurship education	0.50	0.18

NOTES: Returns from fostering entrepreneurship at ages 30 and 40 relative to the returns from fostering entrepreneurship at the beginning of the individuals' careers. Ratio of returns computed using Equation (14). *Subsidy*: $25,000 given to those who participate in incorporated entrepreneurship. *Entrepreneurship Education*: individual-specific signal about incorporated ability given to everyone; intervention characterized by the noise variance of its signals, σ_ν, expressed as a scaled version of the estimated noise variance of trying incorporated entrepreneurship, that is, $\sigma_\nu^2 = \kappa \cdot \sigma_{\eta 4}^2$, where $\kappa = 2.1$.

learning, a subsidy offered later in life may generate more returns because older individuals on the entry margin will have more information than young individuals entering entrepreneurship as a result of the policy.

To explore whether the focus on young entrepreneurship is misplaced, I simulate the policies above (subsidies and education programs for incorporated entrepreneurship) offered at older ages. Concretely, I implement the $25,000 subsidy and the entrepreneurship education program with calibrated information quality ($\kappa = 2.1$) at ages 30 and 40 and compare the returns of these policies relative to the returns the policies yield when offered at the beginning of the individuals' careers (Table 6). The ratio of returns from policy p implemented at age a relative to policy p implemented for the young equals:[56]

$$\frac{(PVI_{p,a} - PVI_{baseline})/Cost_{p,a}}{(PVI_{p,young} - PVI_{baseline})/Cost_{p,young}}. \tag{14}$$

Counterfactual results in Table 7 show that the returns of both policies seem to decrease monotonically with age. This is not necessarily surprising as the amount of years available to reap the benefits of an intervention decreases with age. Consistent with the theory, however, returns from entrepreneurship education fall more rapidly early on. By age 30, the returns from entrepreneurship education have declined 50%, whereas the returns from the subsidy have declined only 14%. Returns from subsidies do not decline as rapidly early on because correlated learning acts as a countervailing force: older individuals entering due to the subsidy have already reduced the uncertainty around their entrepreneurial ability. Better selection notwithstanding, older individuals have less time to reap the benefits of successful experimentation with entrepreneurship; by age 40, the returns from both policies have fallen by at least 70%. Results here show that policymakers trying to foster entrepreneurship are right to focus on the young.[57]

8. CONCLUSION

On the basis of its potential economic benefits, entrepreneurship is the target of policy interventions at different stages in the life cycle. Data show that most people do not attempt entrepreneurship during their careers, and those who attempt it do it after going through several years in paid employment. I study the mechanisms behind these stylized facts thereby exploring the channels through which entrepreneurship policies act. These mechanisms include

[56] The subsidy offered at age $a \in \{30, 40\}$ is the future value of $25,000 at age a. Hence, the cost of the subsidy in present value equals $25,000 multiplied by the number of individuals who participate at age a. Since the entrepreneurship education program is given to everyone, I assume that the present value of program cost does not depend on the age at which it is offered. Therefore, the ratio of program returns when offered at age a relative to program returns when offered at young age conveniently becomes $\frac{PVI_{p,a} - PVI_{baseline}}{PVI_{p,young} - PVI_{baseline}}$.

[57] Counterfactual results (Appendix A.4.9) show that the amount of years remaining in the labor market and the marginal value of new information for uninformed individuals are important mechanisms behind the decline in returns.

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. The article also provides a framework for evaluating entrepreneurship policies. I find that policies that provide information could generate higher returns in present value of lifetime income than subsidies. I also provide evidence, suggesting that policymakers trying to foster entrepreneurship are right to focus on the young.

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 toward 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.

APPENDIX

A.1 *Data Appendix.*

A.1.1 *PSID data.*

This article uses data from the 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, after which the study became biennial. The sample is restricted to white and black men between years 1968 and 1996. Survey data used include occupation, self-employment status, business ownership, incorporation status, labor income, business income, wealth, working hours, completed education, age, race, and marital status.

Potential experience. 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

$$PotentialExperience = Age - CompletedEducation - 6. \tag{A.1}$$

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

$$PotentialExperience = \begin{cases} 0 & if \quad \min PotentialExperience \leq 0 \\ \bar{x} & if \quad \min PotentialExperience = \bar{x} \in \{1, 2, 3\}. \end{cases} \tag{A.2}$$

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 cataloged into one of the salaried occupations.

Occupation. The PSID provides three-digit occupation codes from the 1970 Census of Population that is built using the Alphabetical Index of Industries and Occupations issued June 1971 by the U.S. Department of Commerce and the Bureau of the Census. 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 are coded retroactively 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 article are broad enough and that survey officials get more accurate cataloguing occupations over time, this problem should be minor in the sample.

Not-working status. Individuals are classified as not working 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% 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. Although about 82% of self-employed answer "yes" to this question, less than 8% of salaried workers do.

↪ 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 the 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. Once 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 classified as not working are assigned zero income. All values are in constant 2000 dollars.

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

TABLE A.1
PARAMETERS OF THE WEALTH PROFILE EQUATION

	coeff	se
γ_0	417.78	(85.16)
γ_1	−23.95	(4.40)
γ_2	0.47	(0.06)

NOTES: Wealth in thousands of constant 2000 dollars.

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 article's definition of entrepreneurship, not all years are available for this question. An imputation algorithm is followed 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 are still missing or ambiguous, past and future answers ($t-5, \ldots, t-1, t+1, \ldots, t+5$) from IQ1 and IQ2 are used to assign the incorporated status at t; (iv) all remaining ambiguous observations are imputed as "unincorporated." Out of 2,201 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 the year. 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 (13–15 years of education), college (16 years of education), and 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 in most years in the sample period, I use all wealth observations available (even those outside of the sample period) to construct an individual measure of permanent wealth yielding from the following fixed effects regression:

$$Wealth_{it} = \gamma_0 + \gamma_1 age_{it} + \gamma_2 age_{it}^2 + u_i + \epsilon_{it}. \tag{A.3}$$

Using results in Table A.1, the individual permanent wealth measure is obtained as

$$\omega_i = \hat{\gamma}_0 + \hat{u}_i. \tag{A.4}$$

In estimation, only individuals with at least three wealth data points are included. Figure A.1 shows the age profile of wealth accumulation from pooling all available wealth data.

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

TABLE A.2
INCOME PARAMETERS

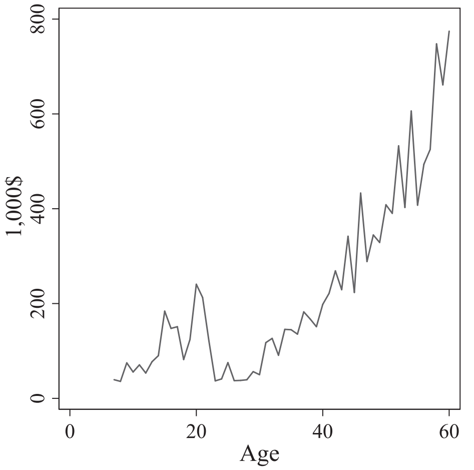
	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Constant	0.927	(0.008)	0.847	(0.021)	1.304	(0.050)	0.730	(0.159)
Black	−0.204	(0.009)	−0.122	(0.015)	−0.180	(0.043)	−0.097	(0.106)
Some College	0.222	(0.009)	0.129	(0.016)	−0.095	(0.044)	0.597	(0.192)
College	0.323	(0.022)	0.456	(0.021)	0.537	(0.109)	0.411	(0.140)
More than College	0.251	(0.026)	0.624	(0.022)	0.934	(0.122)	1.842	(0.325)
Married	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)						

NOTES: Estimated parameters of hourly income in Equation (2) measured in \$10s. This table includes point estimates (coeff) and standard errors (se) corrected for two-stage estimation using subsampling estimation over 100 subsamples without replacement. Returns to experience are estimated as step functions. As an example, $1\{exp_{eu} \geq 3\}$ indicates that the individual has three years or more of unincorporated experience. In blue collar work, experience from both entrepreneurial occupations is pooled: $exp_e = exp_{eu} + exp_{ei}$. Steps' functions were chosen to avoid out of sample return estimates, especially 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.

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;[58] 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.

[58] Relevant data on income and occupation are only collected for household heads.



NOTES: Measured in thousands of dollars of 2000.

FIGURE A1

AVERAGE WEALTH

Individual-year observations satisfying observation from the beginning of their careers: 37,759. (This criterion, and the one for wealth data availability 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. 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, the 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.1.2 *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) that 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 come from the Federal Reserve's Economic Research Data Base and are based on Treasury bills with maturities 1, 2, 3, 5, 7, 10, 20, and 30. Assume that 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

$$b_t \equiv \sum_{s=1}^{\infty} \prod_{i=1}^{s} (1+r_{it})^{-1} = \sum_{s=1}^{30} \prod_{i=1}^{s} (1+r_{it})^{-1} + \prod_{i=1}^{30} (1+r_{it})^{-1} \sum_{s=31}^{\infty} (1+r_{30,t})^{s-30} \tag{A.5}$$

$$= \sum_{s=1}^{30} \prod_{i=1}^{s} (1 + r_{it})^{-1} + \frac{1}{r_{30,t}} \prod_{i=1}^{30} (1 + r_{it})^{-1}.$$

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

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

$$b_t = \sum_{s=1}^{30} \left(\frac{1}{1 + \widetilde{r}_{st}} \right)^{s} + \frac{1}{r_{30,t}} \left(\frac{1}{1 + \widetilde{r}_{30,t}} \right)^{30}. \tag{A.6}$$

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

$$\widetilde{b}_t = \frac{b_t}{deflator2000}. \tag{A.7}$$

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

A.2 *Model Appendix.*

A.2.1 *The updating rules.* Updating rules for similar problems have been previously obtained in the literature (James, 2011). Define the K-dimensional vector ζ_t with characteristic component $\zeta_{\{k\}t}$ and the $K \times K$ diagonal matrix Σ_t with characteristic component $\Sigma_{\{k,k\}t}$ as follows:

$$\zeta_{\{k\}t} = \begin{cases} \zeta_{kt} \text{ if } d_{kt-1} = 1 \\ 0 \text{ otherwise} \end{cases} \qquad \Sigma_{\{k,k\}t} = \begin{cases} 1/\sigma^2_{\eta_k} \text{ if } d_{kt-1} = 1 \\ 0 \text{ otherwise.} \end{cases} \tag{A.8}$$

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

$$\mathbb{E}_t = \left[\mathbb{V}_{t-1}^{-1} + \Sigma_t \right]^{-1} \left[\mathbb{V}_{t-1}^{-1} \mathbb{E}_{t-1} + \Sigma_t \zeta_t \right], \tag{A.9}$$

$$\mathbb{V}_t = \left[\mathbb{V}_{t-1}^{-1} + \Sigma_t \right]^{-1}. \tag{A.10}$$

These updating rules reflect how beliefs change as a function of experience and information. Equations (A.8) and (A.9) imply that the effect of a very noisy signal—high idiosyncratic variance $\sigma^2_{\eta_k}$—on the prior mean is minor. Notably, Equation (A.9) 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.[59] Equation (A.10) 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 that experience is already included in the individual's state, Equation (A.10) implies that $\mathbb{V}_t$ is redundant information.

A.2.2 *Proof of Proposition 1.* PROOF. The proof works by backward induction. Following Margiotta and Miller (2000), and dropping the index i for simplicity, the value function solving his consumption and savings problem at retirement age $T+1$ in present value terms is:

$$V_{T+1}(h_{T+1}, \mathbb{B}_{T+1}, \xi_{T+1}, a_{\tau(T+1)}, b_{\tau(T+1)}) = -\lambda_{\tau(T+1)} b_{\tau(T+1)} \exp\left(\frac{-(\rho\xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right). \tag{A.11}$$

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 $b_{\tau(T+1)}$.[60] 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:

$$-\alpha_{Tk}(h_T)\beta^T \exp\{-\rho c_T - \varepsilon_{Tk}\} - E_T\left[\lambda_{\tau(T+1)} b_{\tau(T+1)} v_{kT+1} \exp\left(\frac{-(\rho\xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right)\middle|\mathbb{B}_T, h_T\right] \tag{A.12}$$

$$s.t.\ \ E_T[\lambda_{\tau(T+1)}\xi_{T+1}|d_{Tk}, h_T, \mathbb{B}_T] + \lambda_{\tau(T)} c_T = \lambda_{\tau(T)}\xi_T.$$

The budget constraint in (A.12) 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 that is simply added to his wealth in Equation (A.11). Following a similar procedure as in Margiotta and Miller (2000, p. 680), the conditional value function of choosing alternative k is obtained as:

$$\begin{aligned} &V_{kT}(h_T, \mathbb{B}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}, \varepsilon_{kT}) = \\ &\quad -\lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} e^{-\varepsilon_{kT}/b_{\tau(T)}} E_T[v_{kT+1}|\mathbb{B}_T]^{1-1/b_{\tau(T)}} \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right). \end{aligned} \tag{A.13}$$

Integrating over ε_T and averaging over the K choices using the ccps yields:

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

[59] The marginal effect of a signal from occupation k on the next period's prior mean of occupation k' equals $(1/\sigma^2_{\eta_k})\mathbb{V}_{\{k,k'\}t}$.

[60] More complex models could have ability and accumulated human capital as determinants of retirement income. 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.

$$\times \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right)$$

$$= -\lambda_{\tau(T)} b_{\tau(T)} \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) A_T(h_T, \mathbb{B}_T), \tag{A.14}$$

where

$$E_\varepsilon[e^{-\varepsilon^*_{kT}/b_{\tau(T)}}] \equiv E_\varepsilon[e^{-\varepsilon_{kT}/b_{\tau(T)}} | d_{kT} = 1]. \tag{A.15}$$

$A_T(h_T, \mathbb{B}_T)$ 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:

$$\begin{aligned} &-\alpha_{kt}(h_t)\beta^t \exp\{-\rho c_t - \varepsilon_{kt}\} \\ &-E_t\left[\lambda_{\tau(t+1)} b_{\tau(t+1)} A_{t+1}(h_{t+1}, \mathbb{B}_{t+1}) v_{kt+1} \exp\left(\frac{-(\rho\xi_{t+1} + a_{\tau(t+1)})}{b_{\tau(t+1)}}\right) \middle| \mathbb{B}_t, h_t, d_{kt} = 1\right] \\ &\quad s.t.\ \ E_t[\lambda_{\tau(t+1)}\xi_{t+1} | d_{kt}, h_t, \mathbb{B}_t] + \lambda_{\tau(t)} c_t = \lambda_{\tau(t)}\xi_t, \end{aligned} \tag{A.16}$$

which yields an equation similar to Equation (A.13):

$$\begin{aligned} &V_{kt}(h_t, \mathbb{B}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}, \varepsilon_{kt}) = \\ &\quad - \lambda_{\tau(t)} b_{\tau(t)} \alpha_{kt}(h_t)^{1/b_{\tau(t)}} e^{-\varepsilon_{kt}/b_{\tau(t)}} E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1}) v_{kt+1} | \mathbb{B}_t, h_t]^{1-1/b_{\tau(t)}} \\ &\quad \times \exp\left(\frac{-(\rho\xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right). \end{aligned} \tag{A.17}$$

The proof is finished by integrating over ε_t and averaging over the K choices using the conditional choices probabilities. ■

A.2.3 *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:

$$\frac{p_{0t}(h_t, \mathbb{B}_t)}{p_{kt}(h_t, \mathbb{B}_t)} = \alpha_{kt}(h_t) E_t\left[\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})}{A_{t+1}(h_t, \mathbb{B}_t)} v_{kt+1} \middle| \mathbb{B}_t, h_t\right]^{b_{\tau(t)}-1}. \tag{A.18}$$

Equation (A.18) 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 (A.18) to write:

$$\begin{aligned} &E_t\big[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1}) v_{kt+1} \big| \mathbb{B}_t, h_t\big]^{1-1/b_{\tau(t)}} \\ &\quad = \alpha_{kt}(h_t)^{-1/b_{\tau(t)}} A_{t+1}(h_t, \mathbb{B}_t)^{1-1/b_{\tau(t)}} \left(\frac{p_{kt}(h_t, \mathbb{B}_t)}{p_{0t}(h_t, \mathbb{B}_t)}\right)^{-1/b_{\tau(t)}}. \end{aligned} \tag{A.19}$$

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

$$E_\varepsilon[e^{-\varepsilon^*_{kt}/b_{\tau(t)}}] = p_{kt}(h_t, \mathbb{B}_t)^{1/b_{\tau(t)}} \Gamma\left(\frac{b_{\tau(t)}+1}{b_{\tau(t)}}\right), \tag{A.20}$$

where $\Gamma(\cdot)$ denotes the complete gamma function. Substitute Equations (A.19) and (A.20) in Equation (8) to obtain:

$$A_t(h_t, \mathbb{B}_t) = p_{0t}(h_t, \mathbb{B}_t)^{1/b_{\tau(t)}} \Gamma\left(\frac{b_{\tau(t)}+1}{b_{\tau(t)}}\right) A_{t+1}(h_t, \mathbb{B}_t)^{1-1/b_{\tau(t)}}. \tag{A.21}$$

Using Equation (A.21), we can write the ratio of indices as:

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})}{A_{t+1}(h_t, \mathbb{B}_t)} = \frac{p_{0t+1}(h_{kt}^{(1)}, \mathbb{B}_{kt}^{(1)})^{1/b_{\tau(t)+1}} A_{t+2}(h_{kt}^{(1)}, \mathbb{B}_{kt}^{(1)})^{1-1/b_{\tau(t)+1}}}{p_{0t+1}(h_{0t}^{(1)}, \mathbb{B}_{0t}^{(1)})^{1/b_{\tau(t)+1}} A_{t+2}(h_{0t}^{(1)}, \mathbb{B}_{0t}^{(1)})^{1-1/b_{\tau(t)+1}}}, \tag{A.22}$$

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 $d \in \{d_{kt}, d_{0t+1}, \ldots, d_{0T}\}$ and define:

$$\phi_t(s) \equiv \frac{1}{b_{\tau(t)+s}} \prod_{r=1}^{s-1} (1 - 1/b_{\tau(t)+r}). \tag{A.23}$$

Iterative substitution of Equation (A.21) in (A.22) up to retirement age yields:

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{B}_{kt+1})}{A_{t+1}(h_t, \mathbb{B}_t)} = \prod_{s=1}^{T-t} \left(\frac{p_{0t+s}(h_{kt}^{(s)}, \mathbb{B}_{kt}^{(s)})}{p_{0t+s}(h_{0t}^{(s)}, \mathbb{B}_{0t}^{(s)})}\right)^{\phi_t(s)}. \tag{A.24}$$

Plugging Equation (A.24) into Equation (A.18) and applying logarithms finish the proof. ■

A.3 *Estimation Appendix.*

A.3.1 *The likelihood.* Instead of 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 Υ be the collection of parameters of the utility function, let Θ be the collection of income parameters, including the variance parameters of the productivity shocks, σ_{η_k}, and let Δ_n be the covariance matrix of the population ability distribution conditional on education level n. Hence, the likelihood of the data—hourly income and choices—for a person with education level n can be written as:

$$\begin{aligned} \mathcal{L}_i = {} & \prod_{t=t_{i0}}^{T_i} \prod_{k=0}^{4} \Pr\left[d_{kit} = 1 | h_{it}, \mathbb{E}_{it}; \Upsilon, \Theta\right]^{d_{kit}} \\ & \times \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_{i0}}^{T_i} \prod_{j=1}^{4} \Pr\left[y_{jit+1} | h_{it}, \tilde{\mu}_j; \Theta\right]^{d_{jit}} \right\} dF_{\mathcal{M}}(\tilde{\mathcal{M}}; \Delta_n). \end{aligned} \tag{A.25}$$

Equation (A.25) displays two characteristics of the model. First, individuals are heterogeneous in their unobserved ability, $\mathcal{M}_i$. Second, instead of 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. Using (A.25), the log likelihood can be separated additively:

$$\ln \mathcal{L}_i = \ln \mathcal{L}_i^d + \ln \mathcal{L}_i^y. \tag{A.26}$$

The first stage of the estimation procedure utilizes the income term of the log likelihood to obtain estimates of Θ and Δ_n. These estimates are used in the second stage to estimate Υ. The scale of Θ, Δ_n, and ρ depends on the units in which income and consumption are measured. Hourly income is expressed in \$10 units and consumption in \$1,000 units. Therefore, converting hourly income into annual income for occupation k in the model entails dividing $\bar{L}_k y_{kt+1}$ by 100.

A.3.2 *First-stage detailed.* Since $\ln \mathcal{L}_i$ is additively separable, $\ln \mathcal{L}_i^y$ is used to consistently estimate Θ and Δ_n using an EM algorithm. To implement the EM algorithm, assume that $\mathcal{M}_i$ is observed for all i. Hence, the income term of the log likelihood for individual i becomes:

$$\ln \mathcal{L}_i^y(\mathcal{M}_i) = \sum_{t=t_{i0}}^{T_i} \sum_{j=0}^{4} d_{jit} \ln \Pr\left[y_{jit+1} | h_{it}, \mu_j; \Theta\right]. \tag{A.27}$$

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 mth iteration:

$$E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i)|\cdot]. \tag{A.28}$$

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

$$\langle \Theta^{m+1},\ \Delta^{m+1} \rangle = \max_{\langle \Theta,\ \Delta \rangle} \sum_i E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i)|\cdot]. \tag{A.29}$$

Expectation step. Using Bayes' rule, the conditional distribution of $\mathcal{M}_i$ for an individual with education level n at the mth iteration based on the observed data is $N(\mathbb{E}_i^m, \mathbb{V}_i^m)$ where[61]

$$\mathbb{E}_i^m = \left((\Delta_n^m)^{-1} + \Psi_i\right)^{-1} \mathbf{W}_i, \tag{A.30}$$

$$\mathbb{V}_i^m = \left((\Delta_n^m)^{-1} + \Psi_i\right)^{-1}, \tag{A.31}$$

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

$$\mathbf{W}_{i\{k\}} = \frac{\sum_{t=1}^{T} d_{kit}\left(y_{kit} - h_{it}'\theta_k\right)}{\sigma_{\eta_k}^{2,m}}, \tag{A.32}$$

and the diagonal components of the square diagonal matrix Ψ_i are

$$\Psi_{i\{k,k\}} = \frac{\sum_{t=1}^{T} d_{kit}}{\sigma_{\eta_k}^{2,m}}. \tag{A.33}$$

[61] Derivations can be found in DeGroot (1970, Ch. 9) and James (2011).

Using μ_{ki} and the distribution of η_{kit} yields:

$$\text{(A.34)} \qquad \log \Pr\left[y_{kit}|h_{it}, \mu_k; \Theta\right] = \log\left(\frac{1}{\sqrt{2\pi\sigma^2_{\eta_k}}}\exp\left\{\frac{-\left(y_{kit} - h'_{it}\theta_k - \mu_{ki}\right)^2}{2\sigma^2_{\eta_k}}\right\}\right)$$
$$= -\frac{1}{2}\log\left(2\pi\sigma^2_{\eta_k}\right) - \frac{1}{2\sigma^2_{\eta_k}}\left(y_{kit} - h'_{it}\theta_k - \mu_{ki}\right)^2.$$

Therefore, the expectation step of the EM algorithm yields:

$$\text{(A.35)} \quad E_m\left[\log \mathcal{L}^y_i\right] = -\sum_{t=1}^{T}\sum_{k=1}^{4} d_{kit}\cdot E_m\left[\frac{1}{2}\log\left(2\pi\sigma^2_{\eta_k}\right) + \frac{1}{2\sigma^2_{\eta_k}}\left(y_{kit} - h'_{it}\theta_k - \mu_{ki}\right)^2\right]$$
$$= -\sum_{t=1}^{T}\sum_{k=1}^{4} d_{kit}\left[\frac{1}{2}\log\left(2\pi\sigma^2_{\eta_k}\right) + \frac{1}{2\sigma^2_{\eta_k}}\left(\mathbb{V}^m_{i\{k,k\}} + \left(y_{kit} - h'_{it}\theta_k - \mathbb{E}^m_{i\{k\}}\right)^2\right)\right],$$

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

Maximization step. Following the expectation step, the maximization step entails maximizing (A.35) to obtain Θ^{m+1}. In fact, each θ^{m+1}_k is given by:

$$\text{(A.36)} \qquad \theta^{m+1}_k = \arg\min_{\theta_k}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{kit}\left(y_{it} - h'_{it}\theta_k - \mathbb{E}^m_{i\{k\}}\right)^2,$$

which yields

$$\text{(A.37)} \qquad \theta^{m+1}_k = (H'W_kH)^{-1}H'W_kY_k,$$

where H is the $[NT \times \#(\theta_k)]$ matrix that stacks together all values of h'_{it}, Y_k is the $[NT \times 1]$ matrix that stacks together all values of $y_{it} - \mathbb{E}^m_{i\{k\}}$, and W_k is the $[NT \times NT]$ diagonal matrix with d_{kit} in its diagonal. Using the first order conditions (FOC) from (A.35) and the estimated values of θ^{m+1}_k, $\sigma^{2,m+1}_{\eta_k}$ has the closed-form solution

$$\text{(A.38)} \qquad \sigma^{2,m+1}_{\eta_k} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} d_{kit}\left(\mathbb{V}^m_{i\{k,k\}} + \left(y_{it} - h'_{it}\theta^{m+1}_k - \mathbb{E}^m_{i\{k\}}\right)^2\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T} d_{kit}}.$$

Consistent with Equation (A.29), the maximization step also includes maximization of the expected value of the log likelihood of $\mathcal{M}_i$ to update population parameter Δ^{m+1}_n. This step is included in the summary below.

Combining the steps. The following is a summary of the EM algorithm:

$\hookrightarrow$ Step 1: Given mth iteration values $\{\theta^m_k, \sigma^{2,m}_{\eta_k}\}_{k\in\{1,\ldots,4\}}$ and $\{\Delta^m_n\}_{s\in\{1,\ldots,4\}}$, solve for $\mathbb{E}^m_i$ and $\mathbb{V}^m_i$ using (A.30) and (A.31).

$\hookrightarrow$ Step 2: Update population parameter Δ_n^{m+1} for education level n as:

$$\Delta_n^{m+1} = \frac{1}{N_n}\sum_{i=1}^{N}\sum_{n=1}^{4}\delta_{in}(\mathbb{V}_i^m + \mathbb{E}_i^m\mathbb{E}_i^{m\prime}), \tag{A.39}$$

where δ_{in} is an indicator of individual i having education level n and $N_n = \sum_i \delta_{in}$. Equation (A.39) 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).

$\hookrightarrow$ Step 3: For each occupation $k > 0$, new iteration values θ_k^{m+1} are obtained using Equation (A.37) and new iteration values $\sigma_{\eta_k}^{2,m+1}$ are obtained using Equation (A.38).

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

$$\left\| \sum_{i=1}^{N} \log \tilde{\mathcal{L}}_i^{y,m+1} - \sum_{i=1}^{N} \log \tilde{\mathcal{L}}_i^{y,m} \right\| < \epsilon, \tag{A.40}$$

where

$$\tilde{\mathcal{L}}_i^{y,m} = \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_{i0}}^{T_i}\prod_{j=1}^{4} \Pr\left[y_{jit+1}|h_{it}, \tilde{\mu}_j; \Theta^m\right]^{d_{jit}} \right\} f_{\mathcal{M}}(\tilde{\mathcal{M}}; \Delta_s^m) \tag{A.41}$$

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

Consistent estimates of individual beliefs are also obtained for use in the second stage. This computation uses the point estimates for Θ and Δ_n, the history of signals received by every individual, Bayes' Rule, and the rational expectations assumption regarding the individual's prior.

A.3.3 *Second-stage detailed.* To estimate Υ, 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:[62]

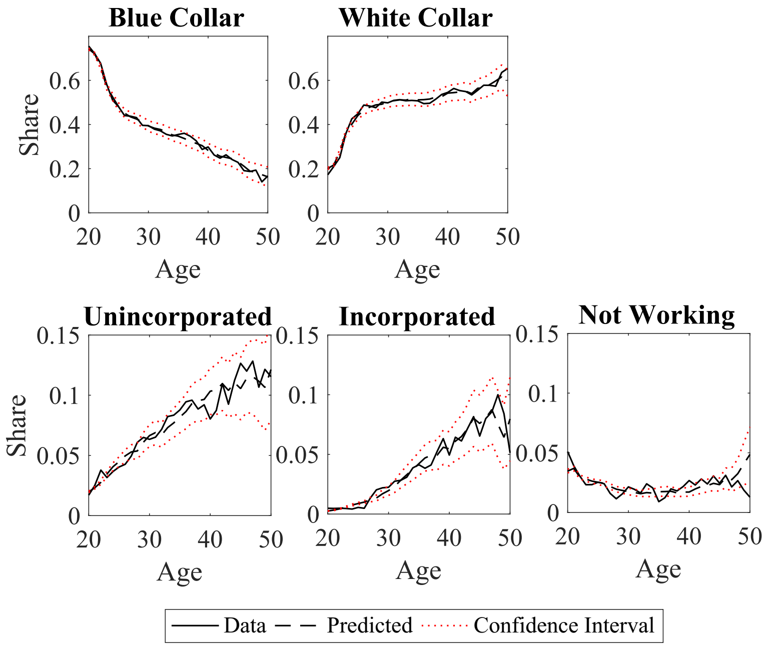
$$p_{kit}(h_{it}, \mathbb{E}_{it}) = \frac{\exp(V_k(h_{it}, \mathbb{E}_{it}))}{1 + \sum_{k'>0}\exp(V_{k'}(h_{it}, \mathbb{E}_{it}))}, \tag{A.42}$$

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

$$V_k(h_{it}, \mathbb{E}_{it}) = -\ln\alpha_{kit}(h_{it}) - (b_{\tau(t)} - 1)\ln E_t\left[v_{kit+1} \prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{0it}^{(s)})} \right)^{\phi_t(s)} \middle| \mathbb{E}_{it}, h_{it} \right]. \tag{A.43}$$

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 ccps estimated from the data (Figure A.2) where beliefs, estimated in the first stage, are treated as data. It entails the following two steps:

[62] Although the bond prices $b_{\tau(t)}$ are part of the state, to save on notation, they are excluded here from the arguments of $p_{kit}(\cdot)$ but included in estimation.



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

FIGURE A2

AVERAGE CONDITIONAL CHOICE PROBABILITIES [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]

1. *Maximization step.* Plug the estimated ccps in Equation (A.42) and maximize the log likelihood of the observed choices.
2. *CCP step.* Use the estimated parameters at the current iteration to solve the model backward 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 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 in 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.

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:

$$\frac{1}{NT}\sum_i\sum_t\sum_{k=0}^{4} d_{kit}\ln p_{kit}(h_{it},\mathbb{E}_{it}). \tag{A.44}$$

The expectation in the expression for $V_k(h_t,\mathbb{E}_t)$ in Equation (A.43) can be written as:

$$E_t\left[v_{kit+1}\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)},\mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)},\mathbb{E}_{0it}^{(s)})}\right)^{\phi(s)}\middle|\mathbb{E}_{it},h_{it}\right] \tag{A.45}$$

$$= \int_{\zeta_k} \left\{ \exp\left(\frac{-\rho \bar{L}_k y_{kit+1}(h_{it})}{b_{\tau(t+1)}} \right) \prod_{s=1}^{T-t} \left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi(s)} \right\} dF_{\zeta_k}(\zeta_k | \mathbb{E}_{it}, h_{it})$$

$$= \int_{\zeta_k} \left\{ \exp\left(\frac{-\rho \bar{L}_k (f_k(h_{it}, \omega_i; \theta_k) + \zeta_k)}{b_{\tau(t+1)}} \right) \prod_{s=1}^{T-t} \left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi(s)} \right\} dF_{\zeta_k}(\zeta_k | \mathbb{E}_{it}, h_{it}).$$

The value of the integral in (A.45) is obtained using Monte Carlo integration. Given a value for ρ, the model becomes a simple logit in the α parameters and the Monte Carlo integral becomes:[63]

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

where

$$A_{kit}(\zeta_k^s) = \prod_{s=1}^{T-t} \left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k^s))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})} \right)^{\phi(s)} \tag{A.47}$$

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

$$\zeta_k^s = \mu_{ki} + \eta_{kit}, \quad \therefore \quad \zeta_k^s \sim N\big(\mathbb{E}_{\{k\}it}, \mathbb{V}_{\{k,k\}it} + \sigma_{\eta_k}^2\big), \tag{A.48}$$

where $\mathbb{E}_{\{k\}it}$ is the kth component of the vector $\mathbb{E}_{it}$ and $\mathbb{V}_{\{k,k\}it}$ is the kth component of the diagonal of the matrix $\mathbb{V}_{it}$. Using the definition of $\alpha_{kt}(h_{it})$ in Equation (5), Equation (A.43) can be rewritten as:

$$V_k(h_{it}, \mathbb{E}_{it}) = -h_{it}' \alpha_k - C_{kit}(\rho), \tag{A.49}$$

where

$$C_{kit}(\rho) = (b_{\tau(t)} - 1) \ln B_{kit}(\rho). \tag{A.50}$$

Equation (A.49) is then substituted into (A.42). In estimation, the log likelihood is maximized conditional on a value of ρ. Search over a discrete set of values of ρ 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 (A.45) more than once for each value of ρ.

CCP step. For a given value of utility parameters, the model is solved backward (Appendix A.4.2) and new model-generated ccps are obtained. These new ccps are fed into the maximization step and new utility parameters are obtained. The iterative process is stopped after five 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

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

process are themselves consistent. In particular, Figure A.2 shows that the initial ccps fed into the iterative process fit the data well.

A.4 *Results Appendix.*

A.4.1 *Parameter estimates.* Table A.6 shows that the estimate of risk aversion in Table A.5—using life cycle income data from paid employment and entrepreneurship—is within the spectrum of results found in the previous literature that uses sources as varied as managerial compensation (Gayle and Miller, 2009; Gayle et al., 2015), insurance markets (Cohen and Einav, 2007; Barseghyan et al., 2013; Handel, 2013), and game shows (Gertner, 1993; Metrick, 1995).

A.4.2 *Solving the model.* As mentioned in Appendix A.3.3, solving the model is required to provide new estimates of the ccps 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 (A.43). Notice that this representation is obtained as a function of the probability of not working in the future conditional on specific choice paths, adjusting for the fact that these paths may not be optimal. 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 $t = T$:

$\hookrightarrow$ *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 (A.43) and the future choice paths described in Proposition 2.[64]

$\hookrightarrow$ *Step 2*. Obtain relevant ccps for period t using Equation (A.42).

$\hookrightarrow$ *Step 3*. Obtain parametric versions, characterized by Ω_t, of the ccps for period t. Noting that only the not working ccps are needed, the parametric version is obtained using a nonlinear regression that minimizes the distance between the model ccps, $p_{0it}(h_{it}, \mathbb{E}_{it})$, and the following parameterization:

$$\exp(X'_{it}\Omega_t)/(1 + \exp(X'_{it}\Omega_t)), \tag{A.51}$$

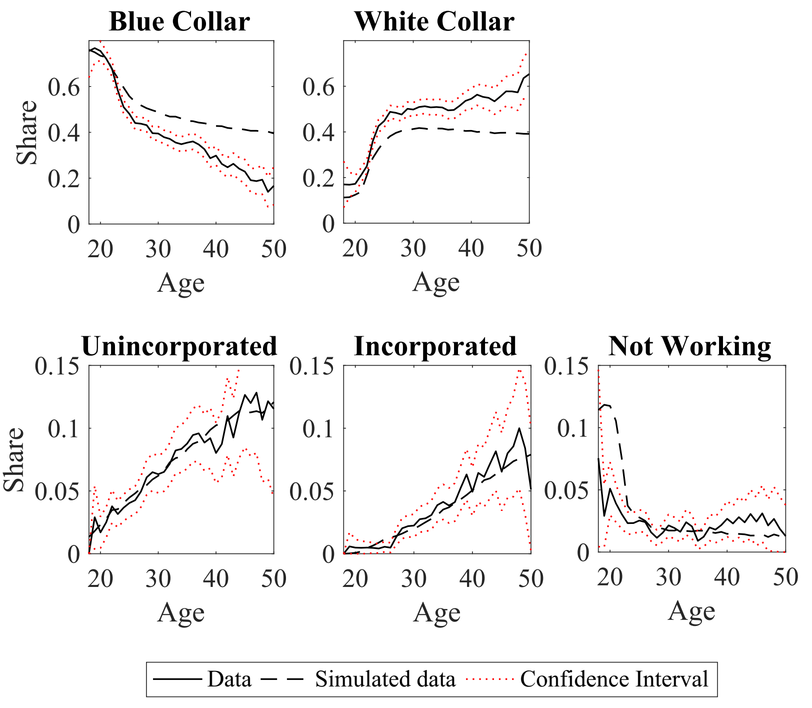
where X_{it} includes multiple interactions of components of the state.

$\hookrightarrow$ *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.

A.4.3 *Model fit.* 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 (A.51) 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 into the labor market, 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 A.3.3. Third, ability for each individual is set at the mean belief conditional on all the information available for him in the data. Using Bayes' rule, this is equal to his beliefs in the last period the individual is observed. To increase precision, only individuals that are observed for at least 10 years are used in the comparison against the data.

[64] 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.



NOTES: Dotted line indicates a 95% confidence interval.

FIGURE A3

SIMULATED VERSUS OBSERVED CHOICES [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]

Comparing simulated data against the sample, Figure A.3 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.[65]

First-entry statistics presented in Table A.7 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.

Table A.8 compares transition matrices from the data and from the simulation. 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 although unincorporated individuals tend to switch in similar percentages into either salaried occupation, incorporated entrepreneurs tend to switch overwhelmingly into white collar work. Table A.9 compares descriptive statistics of occupational spells. Although, consistent with transition results, the model underpredicts spell durations, it performs well in terms of the distribution of occupational spells across occupations.

[65] Figure A.3 is different from Figure 4 since the former, a stricter measure of fit, simulates forward entire paths from an initial state and the latter takes as given the state from the data at every age and simulates current choices.

At the beginning of their careers, the model overpredicts the number of individuals starting as unemployed or blue collar workers, and underpredicts the number starting as white collar workers.

Table A.10 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 overpredicts mean and variance. Notwithstanding this over prediction, the model keeps the relative order of occupations in terms of income variance and average income.

A.4.4 *Certainty equivalent.*

Static. To get a sense of the magnitude of the estimated risk aversion parameter, drop the individual indicator i and 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:

$$\bar{y}_{kt+1} = f_k(h_t;\theta_k) + \mathbb{E}_{t\{k\}}, \tag{A.52}$$

and he considers the variance of his hourly income to be

$$\sigma^2_{kt} = \mathbb{V}_{t\{k,k\}} + \sigma^2_{\eta_k}. \tag{A.53}$$

Therefore, his certainty equivalent at occupation k, y^c_k, solves:

$$-\exp\left\{-\rho\bar{L}_k y^c_k\right\} = -\exp\left\{-\rho\bar{L}_k\bar{y}_{kt+1} + \frac{\rho^2\bar{L}^2_k}{2}\sigma^2_{kt}\right\}, \tag{A.54}$$

which yields

$$y^c_k = \bar{y}_{kt+1} - \frac{\rho\bar{L}_k}{2}\sigma^2_{kt}. \tag{A.55}$$

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

Dynamic. To obtain the dynamic version of the certainty equivalent, use Equation (A.17) to find the quantity y^c_k such that:

$$\begin{aligned} & E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})\exp\left(\frac{-\rho\bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right)|\mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}} \\ &= E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t)\exp\left(\frac{-\rho\bar{L}_k y^c_k}{b_{\tau(t+1)}}\right)|\mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}}. \end{aligned} \tag{A.56}$$

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

$$y^c_k = -\left(\frac{b_{\tau(t+1)}}{\rho\bar{L}_k}\right)\ln\left(\frac{E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})v_{kt+1}|\mathbb{E}_t, h_t\right]}{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t)}\right). \tag{A.57}$$

A.4.5 *Monetary value of entry costs.* The monetary value of entry costs is obtained using Equation (A.17). From Equation (5), one can separate the nonpecuniary costs in two factors: one corresponding to the entry costs, $\alpha^e_{kt}(h_t)$, and the other corresponding to all other nonpecuniary costs. Hence, let $\alpha_{kt}(h_t) = \alpha^o_{kt}(h_t) \times \alpha^e_{kt}(h_t)$. Next, use Equation (5) to find the quantity that should be taken out of annual income in the budget constraint to equalize the conditional value functions. In other words, find the quantity ψ that must be given to the individual to leave him indifferent between (a) receiving ψ and facing entry costs and (b) not receiving ψ but facing no entry costs. The quantity ψ solves:

$$\text{(A.58)} \qquad \alpha^e_{kt}(h_t)^{1/b_{\tau(t)}} E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right) | \mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}}$$
$$= E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho (\bar{L}_k y_{kt+1}(h_t) - \psi)}{b_{\tau(t+1)}}\right) | \mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}},$$

which yields

$$\text{(A.59)} \qquad \psi = \frac{\ln \alpha^e_{kt}(h_t)}{\rho} \frac{b_{\tau(t+1)}}{b_{\tau(t)} - 1}.$$

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

A.4.6 *Alternative regimes.* 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 A.3. Instead of 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.[66] Marital status paths follow the same restriction specified in Appendix A.3.3. 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 A.12.

↪ C1: *No learning-by-doing*. In this counterfactual, individuals receive a fixed 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 during the first 20 years in the labor market for an individual who works exclusively in the occupation. Let $R_k(x)$ be the returns to experience in occupation k for somebody who 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 (A.60) under this counterfactual is given by:

$$\text{(A.60)} \qquad \bar{y}_k = \frac{1}{20} \sum_{x=1}^{20} R_k(x).$$

Individuals under this counterfactual continue to have different returns based on their education, race, marital status, and ability.

[66] 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.

TABLE A.3
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: Covariance matrix of the joint distribution of unobserved ability conditional on education, denoted by Δ_n. This table includes point estimates (coeff) and standard errors (se) corrected for two-stage estimation using subsampling estimation over 100 subsamples without replacement.

TABLE A.4
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: Idiosyncratic hourly income variance in occupation k, denoted by $\sigma^2_{\eta_k}$. This table includes point estimates (coeff) and standard errors (se) corrected for two-stage estimation using subsampling estimation over 100 subsamples without replacement.

TABLE A.5
UTILITY PARAMETERS

	ρ	coeff	se
		0.040	(0.0023)

		High School							
		Blue Collar		White Collar		Unincorporated		Incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Nonpecuniary	*Constant*	−1.818	(0.124)	−0.965	(0.164)	−0.588	(0.182)	−1.551	(0.457)
	Black	0.730	(0.083)	1.182	(0.116)	0.884	(0.157)	1.659	(0.578)
	Married	−0.689	(0.065)	−0.369	(0.087)	−0.544	(0.102)	0.072	(0.446)
Entry cost	*Constant*	−4.778	(1.699)	2.967	(0.616)	5.347	(0.801)	12.963	(0.251)
	$Age/10$	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							
		Blue Collar		White Collar		Unincorporated		Incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Nonpecuniary	*Constant*	−1.985	(0.118)	−1.933	(0.144)	−1.745	(0.182)	−1.258	(0.334)
	Black	0.452	(0.121)	0.839	(0.139)	1.067	(0.231)	1.067	(0.320)
	Married	−0.933	(0.101)	−0.787	(0.108)	−0.673	(0.125)	−0.732	(0.253)
Entry cost	*Constant*	−3.842	(0.419)	3.647	(0.579)	4.978	(0.407)	11.285	(3.445)
	$Age/10$	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							
		Blue Collar		White Collar		Unincorporated		Incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Nonpecuniary	*Constant*	−2.650	(0.181)	−3.614	(0.194)	−3.012	(0.258)	−4.462	(0.337)
	Black	0.184	(0.237)	1.437	(0.298)	1.779	(0.342)	15.975	(3.110)
	Married	−0.210	(0.156)	0.374	(0.184)	0.088	(0.203)	1.078	(0.210)
Entry cost	*Constant*	−3.227	(0.810)	1.147	(0.560)	5.707	(2.343)	7.775	(1.233)
	$Age/10$	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							
		Blue Collar		White Collar		Unincorporated		Incorporated	
	α	coeff	se	coeff	se	coeff	se	coeff	se
Nonpecuniary	*Constant*	−1.075	(0.190)	−1.994	(0.214)	−1.462	(0.271)	−1.005	(0.403)
	Black	−0.674	(0.207)	−0.295	(0.189)	0.470	(0.663)	−0.561	(0.490)
	Married	−0.387	(0.147)	−0.316	(0.145)	−0.713	(0.185)	−0.725	(0.267)
Entry cost	*Constant*	−5.807	(0.606)	−3.242	(0.788)	5.258	(0.619)	9.303	(0.796)
	$Age/10$	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: This table includes point estimates (coeff) of parameters of Equations (4) and (5) and standard errors (se) corrected for two-stage estimation using subsampling estimation over 100 subsamples without replacement. ω_i is defined as the individual's permanent wealth in Section 2 and it is measured in thousands of constant 2000 dollars.

TABLE A.6
ABSOLUTE RISK AVERSION ESTIMATES IN THE LITERATURE

Study	Estimate	Certainty Equivalent
Gayle et al. (2015)	5.01E-07	−626
Gayle and Miller (2009)	5.34E-07	−667
Cohen and Einav (2007)	7.30E-06	−8929
This article	4.00E-05	−33125
Metrick (1995)	6.60E-05	−39518
Gertner (1993)	7.11E-05	−40263
Handel (2013)	1.90E-04	−46352
Barseghyan et al. (2013)	6.40E-04	−48917

NOTES: Certainty equivalent computed using a lottery of equal probability between winning and losing \$50,000. Given the scale of the income data used to estimate risk aversion, I divide the estimates in Gayle et al. (2015) and Gayle and Miller (2009) by 1,000,000 and my own estimate by 1,000 to make them comparable across studies. Cohen and Einav (2007)'s number is the median.

TABLE A.7
FIRST ENTRY: OBSERVED AND SIMULATED

	Data			
	Blue Collar	White Collar	Unincorporated	Incorporated
Tried by age 40	0.65	0.83	0.23	0.11
At first entry				
Age	22.84	24.81	29.57	32.82
exp_{bc}	–	1.99	3.48	2.05
exp_{wc}	1.08	–	3.37	6.58
exp_{eu}	0.03	0.15	–	0.96
exp_{ei}	0.01	0.02	0.22	–
	Model			
	Blue Collar	White Collar	Unincorporated	Incorporated
Tried by age 40	0.77	0.73	0.23	0.09
At first entry				
Age	22.23	25.84	29.98	32.40
exp_{bc}	–	3.32	4.51	3.73
exp_{wc}	0.58	–	3.17	5.53
exp_{eu}	0.08	0.25	–	0.48
exp_{ei}	0.00	0.04	0.11	–

NOTES: Statistics computed using individuals who 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.

$\hookrightarrow$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, income risk is kept constant by setting the value of their idiosyncratic income variance equal to its original value (Table A.4) plus the value of the ability variance (Table A.3). In terms of Equation (2), this amounts to changing the value of the idiosyncratic income variance in occupation k from σ_{η_k} to $\sigma_{\eta_k} + V_{\{k,k\}}$.

$\hookrightarrow$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 0.

$\hookrightarrow$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

TABLE A.8
TRANSITION PATTERNS: OBSERVED AND SIMULATED

Data					
	Blue Collar	White Collar	Unincorporated	Incorporated	Not Working
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
Not working	0.37	0.16	0.04	0.00	0.43
Model					
	Blue Collar	White Collar	Unincorporated	Incorporated	Not Working
Blue collar	0.74	0.18	0.04	0.01	0.03
White collar	0.22	0.71	0.05	0.02	0.01
Unincorporated	0.23	0.21	0.53	0.02	0.01
Incorporated	0.08	0.17	0.03	0.72	0.00
Not working	0.62	0.25	0.05	0.01	0.07

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

TABLE A.9
SPELLS: OBSERVED AND SIMULATED

Data						
	All	Blue Collar	White Collar	Unincorporated	Incorporated	Not Working
Total	4294	1707	1652	453	194	288
Percentage		39.75	38.47	10.55	4.52	6.71
Duration	4.97	5.21	6.03	3.10	3.10	1.63
First		52.06	42.56	2.19	0.27	2.92
Tried		68.73	69.92	20.05	9.03	14.54
Model						
	All	Blue Collar	White Collar	Unincorporated	Incorporated	Not Working
Total	282999	114681	104919	34842	11183	17374
Percentage		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 A.10
INCOME: OBSERVED AND SIMULATED

Data				
	Blue Collar	White Collar	Unincorporated	Incorporated
Mean income	14.14	21.17	21.00	37.48
Variance income	7.94	14.25	22.77	51.17
Model				
	Blue Collar	White Collar	Unincorporated	Incorporated
Mean income	15.21	22.35	23.88	51.04
Variance income	5.83	13.04	27.47	89.57

TABLE A.11
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 (A.60) and the profiles in Table A.2.

TABLE A.12
COMPARISON OF COUNTERFACTUAL REGIMES

Unincorporated

	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 five 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	2.21	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
At first entry									
Ability ($ per hour)	0.49	0.60	5.91	0.91	−0.94	5.67	0.26	0.50	0.16
Belief ($ per hour)	0.38	0.56	–	0.73	0.00	–	0.11	0.40	−0.16
Age	34.1	34.9	32.8	34.9	34.0	32.5	28.0	34.2	34.3
exp_{bc}	6.76	8.80	6.01	9.76	6.47	5.90	0.85	6.80	5.40
exp_{wc}	4.77	3.38	4.25	2.55	5.09	4.04	3.88	4.92	6.67
Overall									
Ability ($ per hour)	3.73	4.21	11.60	4.50	2.09	11.22	2.29	3.69	3.58
College or more	0.55	0.60	0.65	0.63	0.49	0.65	0.44	0.53	0.52

Incorporated

	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 five 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	2.87	5.11	2.91	2.62	5.57	3.39	2.84	3.09
Participation rate at age 40	0.04	0.08	0.10	0.04	0.03	0.14	0.13	0.04	0.08
At first entry									
Ability ($ per hour)	5.54	4.20	17.31	4.95	−2.54	15.22	2.02	6.10	5.00
Belief ($ per hour)	6.36	4.78	–	5.41	0.00	–	2.58	7.09	5.92
Age	38.6	40.9	36.1	40.6	38.7	35.4	29.2	39.4	33.9
exp_{bc}	6.95	12.18	5.98	13.50	7.03	6.07	0.45	7.09	2.70
exp_{wc}	8.14	5.18	6.52	3.97	8.09	5.85	5.18	8.75	7.61
Overall									
Ability ($ per hour)	11.82	9.07	26.99	11.15	2.36	23.65	4.25	11.81	11.70
College or more	0.70	0.64	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. 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 article. Columns *C1–C8* 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, computed over the first 20 years, of somebody who works exclusively in occupation k. *C2* is a full information model where the overall level of initial uncertainty is maintained 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 0. *C4* shuts down correlated learning. *C5* is the full information model without uncertainty. In this counterfactual, all idiosyncratic income variance is eliminated. *C6* eliminates the age profile of entry costs. Entry costs are always those of a 35-year-old person. However, entry costs still vary with permanent wealth and education level. *C7* and *C8* set everyone's permanent wealth at the median and at the 99th percentile, respectively. The same sample of simulated individuals, including their ability vector, is kept constant across all counterfactual regimes.

TABLE A.13
SUMMARY STATISTICS UNDER FULL INFORMATION RELATIVE TO BASELINE

	Blue Collar	White Collar	Unincorporated	Incorporated
Ever tried	0.94	0.97	1.09	1.35
Ever tried in first five years	0.93	0.99	1.34	2.55
Participation rate at age 40	0.91	0.86	1.40	2.33
Spell duration	0.95	0.92	1.30	1.78
Age at first entry	0.99	0.99	0.96	0.93
PVI if ever tried	1.08	1.11	1.28	1.47

NOTES: Each number in the table is the ratio of the given statistic computed under full information relative to the statistic computed in the baseline.

results from diagonalizing the variance–covariance matrix of the distribution of ability.

$\hookrightarrow$C5: *No uncertainty*. In this counterfactual, all income risk is eliminated. Individuals have full information about their ability and face no extra uncertainty resulting from idiosyncratic income variation.

$\hookrightarrow$C6: *Uniform entry costs*. In this counterfactual, individuals of all ages pay the same entry cost, provided that they have the same permanent wealth and education level. This cost equals the one faced by a 35-year-old individual (Table A.5).

$\hookrightarrow$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.

The role of information frictions in all occupations. An advantage of my framework over other papers in the literature with more restricted information structures (e.g., Manso, 2016; Dillon and Stanton, 2017) is that I can evaluate how information frictions affect differently salaried versus entrepreneurial occupations. To do this, I compute the ratio of a number of relevant statistics under full information relative to the baseline (an exercise similar to the one in Figure 12) and present the results in Table A.13.

Results show that the largest responses under full information relative to the baseline occur in entrepreneurial occupations. However, they also reveal that nonnegligible changes in salaried occupations that emerge as individuals become fully informed. Participation and spell duration decrease for salaried occupations, whereas they increase for entrepreneurial occupations: 6% (3%) of individuals who in the baseline engage in blue (white) collar work no longer attempt this occupation under full information, and spell duration in blue (white) collar decreases by 5% (8%). Since individuals can select on ability under full information, it is unsurprising that the PVI also increases for those who ever attempt salaried occupations, although the changes are smaller (between 8% and 11% for salaried occupations, between 28% and 47% for entrepreneurial occupations). Age at first entry into salaried occupations also decreases, although only by 1%.

A.4.7 *Calibrating the quality of an entrepreneurship education program.* Like this article, 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 of 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 κ in Equation (13)—in discrete steps and measuring the corresponding change in the aggregate variance of beliefs relative to its initial value. Using this mapping, I locally interpolate to obtain a value of κ that generates a change in aggregate beliefs equal to the one found in von Graevenitz et al. (2010) (8.24%). This exercise indicates that the information from the program is a little under half the quality of the information in the market ($\kappa = 2.1$).

A.4.8 *Counterfactual policies and entry costs.* Given the reduced-form nature of entry costs in the model, as a function of permanent wealth and age, one may be worried that the conclusions from counterfactual policy experiments (i.e., subsidies and entrepreneurship education) could be very different from those implied by a standard life-cycle framework of consumption and savings with liquidity constrains a lá Evans and Jovanovic (1989). In a standard framework, subsidies increase the individual's initial capital stock making it easier for him to attain the optimal level of capital, and ultimately relaxing the credit constraint as the individual would need less credit. In my model, a subsidy relaxes the entry cost barrier altering the individual's net entrepreneurial income linearly, as a lump sum. In both frameworks, individuals entering as a consequence of the subsidy are those with marginally high (expected) ability but who are constrained by resources. What the reduced-form entry costs in my model fail to capture is the fact that those entering with marginally high ability as a consequence of the subsidy can generate more income with the resources received than those who do not enter—this is because financial capital interacts with ability in the production of net entrepreneurial income in the standard entrepreneurship and credit constraints framework. In that sense, my subsidy counterfactuals are likely to underestimate the benefits from the policy in entry and PVI.

As for entrepreneurship education that alters beliefs, the discrepancy against a model with standard credit constraints could arise from how ability enters the production function of net entrepreneurial income. In the standard model, the optimal level of capital conditional on being an entrepreneur, and hence the occupational choice itself, both depend on the beliefs about ability that are modified by the new information received from the policy. Since there is no optimal level of entrepreneurial financial capital in my model, that channel for the policy to act is shut down. In this model, the entrepreneurship education policy that alters beliefs will cause the set of individuals who are liquidity-constrained to change as their expected ability changes; individuals who now believe that they are high ability are more likely to be constrained if they have low wealth, but they might still enter entrepreneurship given their high-ability belief. This is similar to the example in figure 1 of Evans and Jovanovic (1989). Hence, my results suggesting higher entry and better selection from the entrepreneurship education policy should hold in the model with standard credit constraints. It seems possible that I overestimate the benefits from the policy since I do not constrain entrepreneurial net income as a function of financial capital—entrepreneurial ability enters linearly in the income equation. Hence, new young entrepreneurs with marginally high expected ability entering in response to the policy could appear to produce more relative to the standard model because their production functions are not constrained by financial capital. However, this overestimation of the policy's benefits is counterbalanced by the fact that the financial constraint in my framework, which prevents individuals from fully insuring income risk, mainly affects individuals with high expected ability and high uncertainty around it, the latter being more prevalent among the young.

TABLE A.14

FOSTERING OLD VERSUS YOUNG ENTREPRENEURSHIP: EFFECT OF NUMBER OF REMAINING YEARS

	Truncation at Age	
Policy	$t_0 + T - 30$	$t_0 + T - 40$
Subsidy	0.85	0.64
Entrepreneurship education	0.87	0.57

NOTES: Contribution of the number of years remaining in the labor market to the decreasing returns from fostering entrepreneurship at older ages. Ratio of returns computed using Equation (A.61). *Subsidy*: offers $25,000 to those who participate in incorporated entrepreneurship. *Entrepreneurship Education*: individual-specific signal about incorporated ability given to everyone. Intervention characterized by the noise variance of its signals, σ_ν, expressed as a scaled version of the noise variance of trying incorporated entrepreneurship in reality: $\sigma_\nu^2 = \kappa \cdot \sigma_{\eta_4}^2$, where $\kappa = 2.1$.

A.4.9 *Declining returns of entrepreneurship policies.* Subsection 7.3 showed that the returns from policies fostering entrepreneurship decrease with the age at which they are implemented. Two main mechanisms behind these results are the amount of years remaining in the labor market at the introduction of the policy and the amount of information already received when the policy is introduced. I quantify the contribution of the amount of remaining years in the labor market by computing the following ratio: In the numerator, I put the returns of policy p implemented when young (t_0) but with the number of years in the labor market truncated unexpectedly; and in the denominator, I put the returns from the policy when the number of years in the labor market is not truncated. Concretely, I compute the following ratio of returns:

$$\frac{(PVI^a_{p,t_0} - PVI^a_{baseline})/Cost_{p,t_0}}{(PVI_{p,t_0} - PVI_{baseline})/Cost_{p,t_0}}, \tag{A.61}$$

where the superscript $a \in \{30, 40\}$ in the numerator indicates that the PVI is computed over a truncated interval of years given by $[t_0, t_0 + T - a]$. Comparing results in Table 7 and Table A.14 reveals that the number of years remaining in the labor market explains almost entirely the early decline in the returns of subsidies (a ratio of returns of 0.85 versus 0.86) but only part of the early decline in the returns of entrepreneurship education (a ratio of returns of 0.87 versus 0.50).

Thus, I focus on the entrepreneurship education policy to quantify the role of the quantity of information received prior to the implementation of the policy. To do this, I run four additional counterfactuals. In two of them, I give individuals at the beginning of their careers (t_0) the same amount of information they have on average at age $a \in \{30, 40\}$ under the baseline. In the other two counterfactuals, I do the same but I add the entrepreneurship education policy with calibrated quality ($\kappa = 2.1$) implemented at t_0—hence, individuals receive an extra signal regarding their incorporated entrepreneurial ability. Using the simulated data from these counterfactuals, I compute the following ratio of returns:

$$\frac{(PVI^{\tilde{a}}_{p,t_0} - PVI^{\tilde{a}}_{baseline})/Cost_{p,t_0}}{(PVI_{p,t_0} - PVI_{baseline})/Cost_{p,t_0}}, \tag{A.62}$$

where the superscript $\tilde{a} \in \{30, 40\}$ indicates that individuals at the beginning of their life cycle have the amount of information equivalent to what they would have at age $\tilde{a}$ under the baseline. For age 30 (40), the ratio of returns is 0.83 (0.77). Hence, comparing to results in Table A.14, prior information seems to have a slightly larger effect than number of years remaining in the labor market explaining the early decline in the returns of entrepreneurship education but a weaker effect explaining the decline at older ages.

A.5 *Robustness Appendix.*

A.5.1 *Human capital depreciation.* This appendix specifies and estimates a model with human capital depreciation and replicates some of the counterfactual decompositions in Section 6 to study the role of cross-occupation learning in the presence of human capital depreciation. Relative to the model in Section 3, there are two changes. First, occupation-specific experience x_{kt+1} is split between recent x^R_{kt+1} (accumulated in the last four years) and old x^O_{kt+1} (accumulated five years ago or earlier).[67] Experience evolves as follows:

$$x_{k,t+1} = x_{k,t} + d_{k,t}, \tag{A.63}$$

$$x^R_{k,t+1} = \sum_{r=1}^{4} d_{k,t+1-r}, \tag{A.64}$$

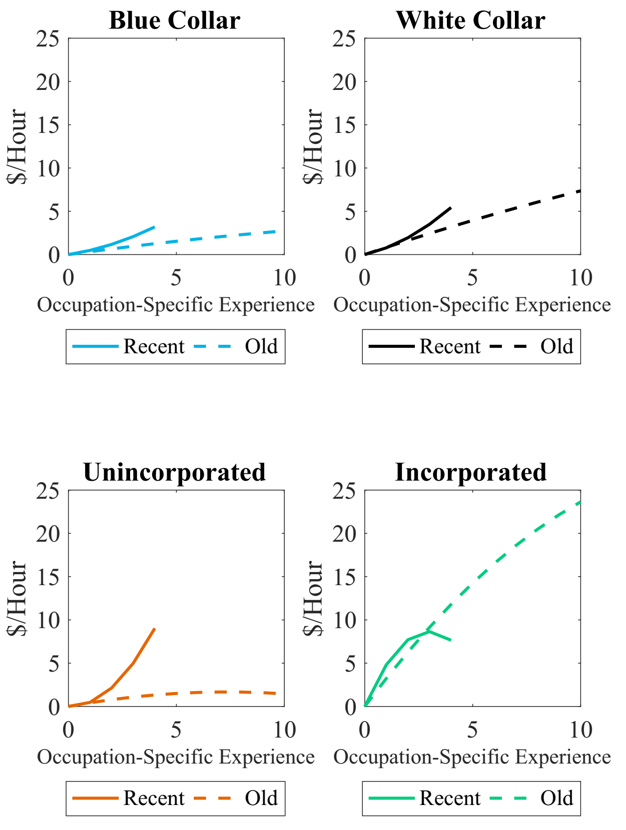
$$x^O_{k,t+1} = x_{k,t+1} - x^R_{k,t+1}. \tag{A.65}$$

Second, for simplicity, the income equation is changed to a quadratic polynomial instead of the flexible piecewise function in the model without depreciation (Figure 6). The depreciation rule increases the number of state variables by 16, 4 (lags) $\times$ 4 (occupations). Since this addition to the size of the state space becomes intractable and I lack enough power to estimate ccps, I assume stochastic transition for the experience vector. In other words, occupation-specific experience (old and new) at t transitions stochastically into $t+1$ conditional on choices at t, d_{kt}, but not conditional on past choices, thereby increasing the state space by only four variables (recent experience). The probabilities for stochastic transitions can be obtained using their sample counterparts. In practice, I use the stochastic transition assumption only when computing ccps (i.e., lags of choices are not included as regressors), and in estimation and simulation, I use the actual lags of recent participation to generate exact transitions. The depreciation rule changes the derivation of the index $A_t(h_t, \mathbb{B}_t)$ in Proposition 1 that captures the future value of beliefs and human capital. Estimation follows the two-stage procedure specified in Appendix A.3.1.

Table A.15 presents the estimates of the income equation. Using these estimates, Figure A.4 shows that for all occupations, the returns to recent own-occupation experience are almost always higher than the returns to old own-occupation experience. Interestingly, the returns to recent and old own-occupation incorporated experience are very similar; this is likely due to the fact that the returns to incorporated experience are highly nonlinear (see Figure 6a) and the second degree polynomial does a poor job approximating the shape of the returns to old experience that are defined over a larger number of years. The estimated returns to cross-occupation experience are less clear-cut. Returns to recent cross-occupation experience dominate returns to old cross-occupation experience for all cross-occupation experience except unincorporated experience. Recent unincorporated entrepreneurship experience is penalized in both white collar work and incorporated entrepreneurship; these results are consistent with those in Figure 6(b).

Figure A.5 compares all first-stage parameters in both models with and without human capital. Returns to experience parameters are not included as they are not directly comparable. The figure shows that all other first-stage parameters in the model with human capital depreciation are very close to the ones estimated in the model without human capital depreciation. Hence, the learning about ability structure of the baseline model remains virtually unchanged. Figure A.6 presents a similar comparison for all second-stage parameters. Utility parameters change slightly

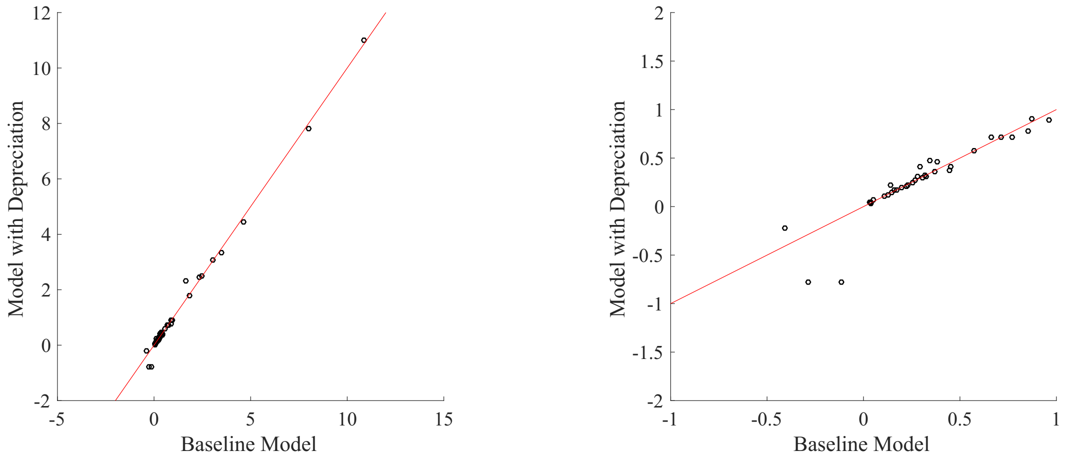
[67] Studying women's labor supply, Altuğ and Miller (1998) find that up to six lags of participation are significant in a wage equation. I let four lags to determine recent experience for tractability.



NOTES: Returns to recent and old own-occupation experience implied by estimates in Table A.15.

FIGURE A4

RETURNS TO OWN-OCCUPATION EXPERIENCE [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]



(a) All parameters (b) Parameters between -1 and 1

NOTES: First-stage parameters except returns to experience parameters (Table A.15). The latter are not comparable since the returns profile is different in the model with human capital depreciation. Red line is the 45° line.

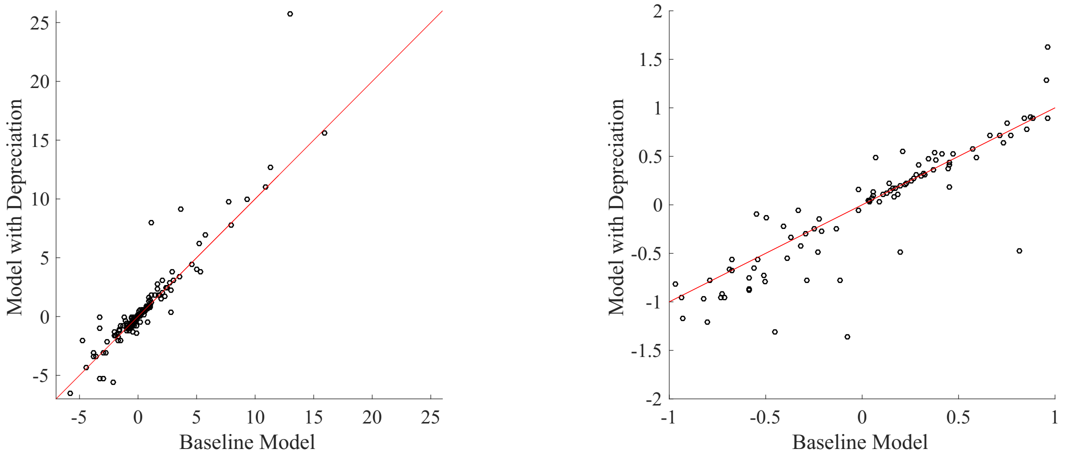
FIGURE A5

COMPARISON OF FIRST-STAGE PARAMETERS [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]

TABLE A.15
INCOME PARAMETERS WITH DEPRECIATION

	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Constant	0.931	(0.007)	0.719	(0.014)	0.962	(0.061)	0.404	(0.213)
Black	−0.200	(0.008)	−0.097	(0.014)	−0.237	(0.051)	−0.071	(0.221)
Some College	0.221	(0.008)	0.138	(0.013)	−0.062	(0.041)	0.941	(0.161)
College	0.337	(0.017)	0.500	(0.017)	0.607	(0.079)	0.615	(0.141)
More than College	0.239	(0.022)	0.688	(0.019)	1.012	(0.077)	1.839	(0.300)
Married	0.048	(0.004)	0.215	(0.008)	0.031	(0.035)	0.717	(0.107)
exp_{bc}^{R}	0.038	(0.004)	0.090	(0.004)	0.094	(0.017)	0.070	(0.072)
$(exp_{bc}^{R})^2$	0.011	(0.001)						
exp_{bc}^{O}	0.034	(0.001)	0.003	(0.001)	−0.014	(0.004)	−0.097	(0.018)
$(exp_{bc}^{O})^2$	−0.001	(6.302)						
exp_{wc}^{R}	0.086	(0.002)	0.058	(0.008)	0.078	(0.017)	0.049	(0.061)
$(exp_{wc}^{R})^2$			0.019	(0.001)				
exp_{wc}^{O}	0.027	(0.001)	0.084	(0.001)	0.060	(0.005)	0.051	(0.011)
$(exp_{wc}^{O})^2$			−0.001	(9.084)				
exp_{eu}^{R}			−0.034	(0.009)	−0.012	(0.032)	−0.012	(0.072)
$(exp_{eu}^{R})^2$					0.060	(0.007)		
exp_{eu}^{O}			0.063	(0.004)	0.046	(0.011)	0.025	(0.029)
$(exp_{eu}^{O})^2$					−0.003	(0.000)		
exp_{ei}^{R}			0.612	(0.012)	0.254	(0.033)	0.580	(0.096)
$(exp_{ei}^{R})^2$							−0.097	(0.021)
exp_{ei}^{O}			0.400	(0.007)	0.165	(0.016)	0.334	(0.034)
$(exp_{ei}^{O})^2$							−0.010	(0.002)
exp_{e}^{R}	0.041	(0.005)						
exp_{e}^{O}	0.002	(0.003)						

NOTES: Hourly income measured in \$10s. This table includes point estimates (coeff) and standard errors (se) not corrected for two-stage estimation. Estimated parameters of Equation (2) under depreciation model. Returns to experience are estimated as polynomials of recent (exp_k^R) and old (exp_k^O) experience. In blue collar work, experience from both entrepreneurial occupations is pooled: $exp_e^r = exp_{eu}^r + exp_{ei}^r$ for $r = R, O$.



(a) All parameters (b) Parameters between -1 and 1

NOTES: All utility parameters included in the comparison. Red line is the 45º line.

FIGURE A6

COMPARISON OF SECOND-STAGE PARAMETERS [COLOR FIGURE CAN BE VIEWED AT WILEYONLINELIBRARY.COM]

TABLE A.16
CROSS-OCCUPATION LEARNING WITH HUMAN CAPITAL DEPRECIATION

Unincorporated

	Without Depreciation		With Depreciation	
	C3	C4	C3	C4
Ever tried	0.92	1.04	1.12	1.63
Ever tried in first five years	0.77	1.02	0.93	2.44
PVI if ever tried	0.95	0.93	0.87	1.10

Incorporated

	Without Depreciation		With Depreciation	
	C3	C4	C3	C4
Ever tried	0.96	0.90	0.99	0.60
Ever tried in first five years	0.60	0.91	0.10	0.38
PVI if ever tried	0.82	0.76	0.46	0.64

NOTES: Ratio of the average of several summary statistics across alternative regimes relative to the average under their respective baseline (i.e., without or with human capital depreciation). *PVI* stands for the present value of income in thousands of dollars. This average is computed only over those who tried the occupation. *Columns*: *C3* sets the cross-occupation returns to experience to 0. *C4* shuts down correlated learning. The first (last) two columns correspond to counterfactuals using the model without (with) human capital depreciation. The same sample of simulated individuals, including their ability vector, is kept constant across all counterfactual regimes.

more than the first-stage parameters but remain closely around those estimated in the model without human capital depreciation. The risk aversion estimate, which is estimated in a grid, is identical at 0.040.

Human capital depreciation could eliminate the role of cross-occupation learning by doing because all cross-occupation experience by definition loses value over time after a switch in occupations. It also decreases the incentives to change occupations after fine-tuning beliefs, potentially undermining the role of cross-occupation learning about ability. To explore these possibilities, I simulate data from the model with depreciation and from two counterfactual versions of this model in which cross-occupation learning by doing and cross-occupation learning about ability are shut down, respectively. These counterfactuals correspond to models C3 and C4 in Appendix A.4.6, except now they include human capital depreciation. Table A.16 presents the results from this exercise; each entry represents the ratio of the respective statistic (rows) in each counterfactual (columns) relative to the respective baseline (with or without human capital). For instance, focusing on the *Incorporated* panel of Table A.16, the first row indicates that in the model without human capital depreciation, the share of individuals who attempt incorporated entrepreneurship when there is no cross-occupation learning by doing (learning about ability) is 0.96 (0.90) times the share of individuals who attempt incorporated entrepreneurship in the baseline; in the model with human capital depreciation, the ratio is 0.99 (0.60). The second row indicates that young entrepreneurship falls more dramatically in the model with depreciation. The PVI in the last row also decreases more in the model with depreciation. Hence, the role of cross-occupation learning by doing and cross-occupation learning about ability, and hence the value of this type of learning, is not eliminated in the presence of human capital depreciation.

A.5.2 *Current wealth.* In this appendix, I explore in reduced form the relationship between current wealth and the decision to switch into entrepreneurship comparing the role of current wealth with that of other variables included in the model. I pool all observations with wealth data available coming from years 1984, 1989, and 1994. Table A.17 shows the results of separate logit regressions for switching into unincorporated and incorporated entrepreneurship from paid

TABLE A.17
LOGIT OF SWITCHING TO ENTREPRENEURSHIP IN $t+1$

	Unincorporated		Incorporated	
Variable	coeff	se	coeff	se
Current wealth				
2nd Quartile	−0.954	(0.398)	0.688	(0.711)
3rd Quartile	−0.305	(0.371)	0.218	(0.795)
4th Quartile	−0.37	(0.46)	0.886	(0.823)
Experience variables				
exp_{bc}	0.004	(0.151)	−0.215	(0.262)
exp_{bc}^2	−0.002	(0.005)	0.004	(0.01)
exp_{wc}	−0.092	(0.141)	−0.060	(0.205)
exp_{wc}^2	0.002	(0.004)	0.001	(0.004)
exp_{eu}	0.884	(0.249)	0.004	(0.823)
exp_{eu}^2	−0.086	(0.029)	−0.103	(0.147)
exp_{ei}	1.379	(0.465)	2.123	(0.529)
exp_{ei}^2	−0.144	(0.067)	−0.230	(0.067)
Beliefs variables				
$\mathbb{E}_{bc}$	−0.41	(1.146)	−1.529	(1.27)
$\mathbb{E}_{wc}$	−0.075	(1.569)	−0.553	(1.594)
$\mathbb{E}_{eu}$	0.22	(0.338)	0.622	(0.655)
$\mathbb{E}_{ei}$	−0.001	(0.549)	0.464	(0.588)
Observations	2,789		2,753	
Pseudo-R^2	0.1345		0.2070	

NOTES: This table includes point estimates (coeff) and standard errors (se). Dependent variable in regressions is whether individual becomes an unincorporated entrepreneur in $t+1$ and whether he becomes an incorporated entrepreneur in $t+1$, respectively. Independent variables are pooled from years 1984, 1989, and 1994. Other independent variables included in the regression but omitted in the table are age, race, education level indicators, and whether the individual is currently in a white collar occupation. Forty-seven percent of observation are in blue collar work in t, and the rest are in white collar. Also 2.4% of those became unincorporated entrepreneurs in $t+1$ and 1.1% became incorporated entrepreneurs in $t+1$. Beliefs variables are obtained from the first-stage estimates of the baseline model (Appendix A.4.1). When estimating the unincorporated regression, the observations of individuals who became incorporated entrepreneurs are excluded. Similarly, when estimating the incorporated regression, observations of individuals who became unincorporated entrepreneurs are excluded.

TABLE A.18
MARGINAL EFFECTS

	Unincorporated		Incorporated	
Variable	mfx	p-Value	mfx	p-Value
Current Wealth: 1*st* to 4*th*	−0.009	0.408	0.008	0.262
exp_{eu}, 0 to 1	0.020	0.004	−0.001	0.886
exp_{ei}, 0 to 1	0.044	0.051	0.040	0.037
$\mathbb{E}_{eu}$	0.002	0.539	0.003	0.412
$\mathbb{E}_{ei}$	0.000	0.999	0.005	0.532

NOTES: Marginal effects (mfx) implied by Logit models in Table A.17. Marginal effects computed as the difference in average switching probability from varying the relevant variable for all individuals while keeping all other variables' values unchanged. For instance: row "Current wealth: 1*st* to 4*th*" evaluates the change in the probability of switching from moving from the first to the fourth quartiles of current wealth; row "exp_{ei} (0 to 1)" evaluates the change from increasing entrepreneurial incorporated experience in one unit from having no prior experience; all belief variables' rows evaluate the change from increasing the belief variable one standard deviation from the variable mean. The *p-value* column evaluates whether the marginal change is statistically significant.

employment. The results suggest that lower (higher) current wealth is associated with higher switching into unincorporated (incorporated) entrepreneurship although all wealth coefficients are insignificant.

To get a sense of the magnitude of the associations, Table A.18 presents the marginal effects of a subset of the regressors. For unincorporated entrepreneurship, the decrease in the

TABLE A.19
LOGIT OF SWITCHING TO ENTREPRENEURSHIP IN 1985

Variable	coeff	se	variable	coeff	se
ESHI variables			*Beliefs Variables*		
$ESHI \cdot 1\{d_{bc}=1\}$	−1.392	0.510	$\mathbb{E}_{bc}$	1.663	1.461
$ESHI \cdot 1\{d_{wc}=1\}$	−0.575	0.633	$\mathbb{E}_{wc}$	−2.803	2.160
			$\mathbb{E}_{eu}$	0.109	0.888
Demographic variables			$\mathbb{E}_{ei}$	1.130	0.654
Age	−0.004	0.150	$\mathbb{V}_{bc,bc}$	17.393	10.669
Black	0.087	0.423	$\mathbb{V}_{bc,wc}$	0.067	18.774
Some College	4.494	2.198	$\mathbb{V}_{bc,eu}$	10.788	12.890
College	3.906	2.919	$\mathbb{V}_{bc,ei}$	−14.874	8.214
More than College	7.139	4.175	$\mathbb{V}_{wc,wc}$	−22.965	19.340
			$\mathbb{V}_{wc,eu}$	−2.450	17.593
Experience variables			$\mathbb{V}_{wc,ei}$	17.659	11.660
$1\{d_{wc}=1\}$	−0.539	0.707	$\mathbb{V}_{eu,eu}$	−0.338	1.293
exp_{bc}	0.158	0.301	$\mathbb{V}_{eu,ei}$	0.190	5.728
exp^2_{bc}	−0.010	0.016	$\mathbb{V}_{ei,ei}$	−2.003	1.060
exp_{wc}	−0.367	0.262			
exp^2_{wc}	0.012	0.013	*Constant*	−3.386	3.288
exp_e	1.570	0.477			
exp^2_e	−0.245	0.089			

NOTES: This table includes point estimates (coeff) and standard errors (se). Dependent variable is whether individual becomes an entrepreneur (either incorporated or unincorporated) in 1985. Independent variables are all from 1984. The sample includes 923 observations, 49.5% are in blue collar work in 1984, and the rest are in white collar. Also 4.2% of those become entrepreneurs (either incorporated or unincorporated) in 1985. *ESHI* stands for employer-sponsored health insurance. Entrepreneurial experience exp_e is the sum of experience in both types of entrepreneurship $exp_e = exp_{eu} + exp_{ei}$. Beliefs variables are obtained from the first-stage estimates of the baseline model (Appendix A.4.1).

probability of switching yielding from going from the bottom to the top quartile of current wealth (−0.009) is 46% the magnitude of the increase in probability from a year of prior experience in unincorporated entrepreneurship (0.020) and four times as large as the increase from a movement of one standard deviation in beliefs about unincorporated ability (0.002). For incorporated entrepreneurship, the increase in the probability of switching resulting from going from the bottom to the top quartile of current wealth (0.008) is 21% the magnitude of the increase in probability from a year of prior experience in incorporated entrepreneurship (0.040) and 74% higher than the increase from a movement of one standard deviation in beliefs about incorporated ability (0.005).

A.5.3 *Employer-sponsored health insurance.* In this appendix, I explore in reduced form the link between ESHI and the probability of switching into self-employment. I use data from 1984, the only year in the PSID that has information on ESHI during my sample period. Individuals' answer, the following question: "Does your employer pay for any medical, surgical, or hospital insurance that covers any illness or injury that might happen to you when you are not at work?" Merging the answers to this question with my sample yields 923 observations of individuals in paid employment in 1984. Using this subsample, I estimate a logit of whether individuals switch to self-employment in 1985. Given the small number of observations, the dependent variable pools both types of self-employment, incorporated and unincorporated.

Results in Table A.19 suggest that ESHI is associated with a lower probability of switching into entrepreneurship, although the coefficient is only statistically significant for individuals who are currently engaged in blue collar work. Table A.19 reveals that the incentives to stay in paid employment provided by ESHI may be stronger in blue collar work. To compare the size of the effect of ESHI to that of other variables included in the structural model, Table A.20 presents the marginal effects. Comparing the effect of ESHI against the effects of learning about ability variables, Table A.20 indicates that the size of the decrease in the switching probability

TABLE A.20
MARGINAL EFFECTS

Variable	mfx	p-Value	Variable	mfx	p-Value
ESHI variables		*Beliefs variables*			
$ESHI \cdot 1\{d_{bc,t} = 1\}$	−0.060	(0.018)	$\mathbb{E}_{eu}$	0.002	(0.904)
$ESHI \cdot 1\{d_{wc,t} = 1\}$	−0.022	(0.421)	$\mathbb{V}_{eu,eu}$	−0.015	(0.785)
			$\mathbb{E}_{ei}$	0.056	(0.238)
Experience variables		$\mathbb{V}_{ei,ei}$	−0.158	(0.000)	
$1\{d_{wc} = 1\}$	−0.004	(0.854)			
exp_{bc}, 0 to 1	0.005	(0.535)	*Demographic variables*		
exp_{bc}, 1 to 2	0.005	(0.606)	*Black*	0.003	(0.840)
exp_{wc}, 0 to 1	−0.022	(0.386)	*Some College*	0.219	(0.109)
exp_{wc}, 1 to 2	−0.016	(0.344)	*College*	0.172	(0.376)
exp_{e}, 0 to 1	0.074	(0.022)	*More than College*	0.470	(0.136)
exp_{e}, 1 to 2	0.095	(0.097)			

NOTES: Marginal effects (mfx) implied by logit model in Table A.19. Marginal effects computed as the difference in average switching probability from varying the relevant variable for all individuals while keeping all other variables' values unchanged. For instance: row "$ESHI \cdot 1\{d_{bc} = 1\}$" evaluates the change in the probability of switching from not having ESHI and working in blue collar to having ESHI and working in blue collar; row "exp_e (1 to 2)" evaluates the change from increasing entrepreneurial experience in one unit from having 1 year of prior entrepreneurial experience; row *somecollege* evaluates the change from increasing education from the base category (high school) to some college; all belief variables' rows evaluate the change from increasing the belief variable one standard deviation from the variable mean. The *p-value* column evaluates whether the marginal change is statistically significant.

from having ESHI while working in blue collar work (0.060) is comparable to the gain yielding from increasing the mean belief of incorporated entrepreneurial ability $\mathbb{E}_{ei}$ by one standard deviation (0.060), but it is only 38% of the decrease in probability generated by increasing the variance belief of incorporated entrepreneurial ability $\mathbb{V}_{ei,ei}$ by one standard deviation (0.158). Comparing the effect of ESHI against the effects of learning by doing variables, Table A.20 indicates that the size of the decrease (0.060) is about 81% of the size of the increase from having one year of prior entrepreneurial experience (0.074). The size of the decrease in the switching probability from having ESHI while working in white collar (0.022) is about 37% of the size of the decrease while working in blue collar (0.060). Overall, these results suggest that the estimated nonincome benefits from paid employment (especially those from blue collar work) may contain benefits associated with ESHI.

A.5.4 *Heterogeneous risk aversion.* Individuals in the model have homogeneous risk aversion. This assumption could affect the counterfactual results depending on which individuals are responsive to policies. For instance, a policy such as entrepreneurship education that shifts beliefs can work through at least two selection mechanisms: ability and risk aversion. On the one hand, better beliefs resulting from the policy are correlated with better ability, encouraging more able individuals to join and deterring less able ones. On the other hand, information signals provided by the policy decrease the variance of beliefs encouraging more risk-averse individuals to join independent of their ability.

To assess the impact of assuming homogeneous risk aversion, I solved the model taking the estimated parameters as given but assuming that risk aversion ρ is i.i.d. $\sim$ log Normal $(\mu_{\ln\rho}, \sigma^2_{\ln\rho})$ across individuals with $\sigma^2_{\ln\rho} = 1$. Under these assumptions:

$$E[\rho] = \exp\{\mu_{\ln\rho} + 0.5\}. \tag{A.66}$$

Hence, I set $\mu_{\ln\rho} = \ln\hat{\rho} - \frac{1}{2}$, where $\hat{\rho}$ is the point estimate of ρ in Table A.5. I solve the model backward, obtain new ccps that include individual-specific risk aversion, and simulate forward the behavior of individuals drawing their risk aversion from the distribution above. I simulate their behavior under the baseline and the entrepreneurship education policy.

TABLE A.21
ENTREPRENEURSHIP EDUCATION AND RISK AVERSION

	Homogeneous		Heterogeneuos	
	Baseline	Program	Baseline	Program
Young entrepreneurs				
Tried in first five years	0.020	0.082	0.014	0.063
Mean belief ($ per hour) at first entry	3.7	48.4	4.3	59.0
Mean ability ($ per hour) at first entry	5.0	10.7	5.6	5.9
Bias (beliefabiliity) at first entry	−1.3	37.7	−1.3	53.1
Overall				
Tried	0.15	0.24	0.11	0.18
Participation rate at age 40	0.04	0.11	0.03	0.07
PVI ($1000s)	757	983	678	896
Mean belief ($ per hour) at first entry	6.4	26.5	5.3	32.5
Mean ability ($ per hour) at first entry	5.5	6.6	4.7	4.0
Bias (beliefabiliity) at first entry	0.9	19.9	0.6	28.5
Mean ability ($ per hour)	11.8	13.4	10.8	10.9
Mean risk aversion	0.040	0.040	0.034	0.051
All individuals				
PVI ($1000s)	508	581	469	516

NOTES: Baseline and entrepreneurship education program under homogeneous and heterogeneous risk aversion. The entrepreneurship education program is identical to that in Subsection 7.2 where the quality of the signal provided is the same as that of the signal received from becoming an incorporated entrepreneur for one period. Numbers in the "Homogeneous" columns are identical to those in the first (*Baseline*) and last (*Entrepreneurship Education*, noise variance scale = 1) columns of Table 6.

Table A.21 presents the results from this exercise and also includes the results using the estimated model with homogeneous risk aversion (Table 6). In general, the entrepreneurship education program has very similar results with homogeneous or heterogeneous risk aversion. Relative to the baseline with homogeneous (heterogeneous) risk aversion, as a consequence of the entrepreneurship education program, the share of young entrepreneurs is 4.1 (4.6) times as high, the proportion of individuals who ever try entrepreneurship grows by 58% (62%), the participation rate at age 40 is 2.6 (2.7) times as high, the PVI for people who try entrepreneurship increases by 30% (32%), and the PVI for all individual increases by 14% (10%).

Mean ability and mean belief of all first-time entrepreneurs constitute the main difference in the effects of the policy: with homogeneous risk aversion, their mean ability increases 18% and their mean belief becomes 4.2 times as large as baseline; and with heterogeneous risk aversion, their mean ability decreases 14% and their mean belief becomes 6.2 times as large as baseline. Table A.21 suggests that with heterogeneous risk aversion, the policy does relax the risk aversion threshold for participation as average risk aversion of all entrepreneurs increases by 50%. New risk-averse entrepreneurs tend to require a higher bias in ability to enter, which explains the decrease in ability at first entry due to the policy. Over time the selection mechanism weeds out entrepreneurs with highly biased beliefs, leaving the average ability of entrepreneurs mostly unchanged at around 10.9, which is reflected in the PVI of those who try entrepreneurship. Therefore, although the risk aversion mechanism does seem to play a small role when evaluating the policy, the effects of the policy obtained in the article remain mainly unchanged.

REFERENCES

AGUIRREGABIRIA, V., AND P. MIRA, "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica* 70 (2002), 1519–43.

ALTUĞ, S., AND P. LABADIE, *Dynamic Choice and Asset Markets* (San Diego, CA: Academic Press, 1994).

———, and R. A. Miller, "The Effect of Work Experience on Female Wages and Labour Supply," *Review of Economic Studies* 65 (1998), 45–85.

Amemiya, T., *Advanced Econometrics* (Cambridge, MA: Harvard University Press, 1985).

Anderson, T., and I. Olkin, "Maximum-Likelihood Estimation of the Parameters of a Multivariate Normal Distribution," *Linear Algebra and its Applications* 70 (1985), 147–71.

Antonovics, K., and L. Golan, "Experimentation and Job Choice," *Journal of Labor Economics* 30 (2012), 333–66.

Arcidiacono, P., and R. A. Miller, "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica* 79 (2011), 1823–67.

———, and ———, "Identifying Dynamic Discrete Choice Models off Short Panels," *Journal of Econometrics*, forthcoming.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum, "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review* 103 (2013), 2499–529.

Buera, F. J., "A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence," *Annals of Finance* 5 (2009), 443–64.

Camerer, C., and D. Lovallo, "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review* 89 (1999), 306–18.

Cohen, A., and L. Einav, "Estimating Risk Preferences from Deductible Choice," *American Economic Review* 97 (2007), 745–88.

DeGroot, M. H., *Optimal Statistical Decisions* (New York, NY: McGraw Hill, 1970).

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. Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *The Journal of Political Economy* 97 (1989), 808–27.

Fairlie, R. W., K. Kapurb, and S. Gatesc, "Is Employer-Based Health Insurance a Barrier to Entrepreneurship?," *Journal of Health Economics* 30 (2011), 146–62.

Gayle, G., L. Golan, and R. A. Miller, "Promotion Turnover and the Compensation in the Executive Labor Market," *Econometrica* 83 (2015), 2293–369.

———, and R. A. Miller, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?," *American Economic Review* 99 (2009), 1740–69.

Gertner, R., "Game Shows and Economic Behavior: Risk-Taking on 'Card Sharks'," *The Quarterly Journal of Economics* 108 (1993), 507–21.

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* (Minneapolis, MN: University of Minnesota Press, 1987), 3–25.

Hahn, Y., and H.-S. Yang, "Do Work Decisions among Young Adults Respond to Extended Dependent Coverage?," *Industrial and Labor Relations Review* 69 (2016), 737–71.

Hall, R. E., and S. E. Woodward, "The Burden of the Nondiversifiable Risk of Entrepreneurship," *American Economic Review* 100 (2010), 1163–94.

Hamilton, B. H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment," *The Journal of Political Economy* 108 (2000), 604–31.

———, N. Pande, and N. W. Papageorge, "The Right Stuff? Personality and Entrepreneurship," *Quantitative Economics* 10 (2019), 643–91.

Handel, B. R. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review* 103 (December 2013), 2643–82.

Heckman, J. J., and B. E. Honoré, "The Empirical Content of the Roy Model," *Econometrica* 58 (1990), 1121–49.

Holtz-Eakin, D., J. R. Penrod, and H. S. Rosen, "Health Insurance and the Supply of Entrepreneurs," *Journal of Public Economics* 62 (1996), 209–35.

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), 319–47.

———, and B. W. Pugsley, "Wealth, Tastes, and Entrepreneurial Choice," in J. Haltiwanger, E. Hurst, J. Miranda, and A. Schoar, eds., *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges* (Chicago: University of Chicago Press, 2017), 111–51.

Iyigun, M. F., and A. L. Owen, "Risk, Entrepreneurship, and Human-Capital Accumulation," *American Economic Review* 88 (1998), 454–7.

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), 972–90.

———, AND Y. NYARKO, "Learning by Doing and the Choice of Technology," *Econometrica* 64 (1996), 1299–310.

KEANE, M. P., AND K. I. WOLPIN, "The Career Decisions of Young Men," *The Journal of Political Economy* 105 (1997), 473–522.

KREIDER, R. M., AND R. ELLIS, "Number, Timing, and Duration of Marriages and Divorces: 2009," Technical Report, U.S. Census Bureau, 2011.

LAFONTAINE, F., AND K. SHAW, "Serial Entrepreneurship: Learning by Doing?," *Journal of Labor Economics* 34 (2016), S217–54.

LAZEAR, E. P., "Entrepreneurship," *Journal of Labor Economics* 23 (2005), 649–80.

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.

———, AND ———, "Selection into Entrepreneurship and Self-Employment," NBER Working Paper 25350, 2018.

LINDQUIST, M. J., J. SOL, AND M. VAN PRAAG, "Why Do Entrepreneurial Parents Have Entrepreneurial Children?," *Journal of Labor Economics* 33 (2015), 269–96.

MACURDY, T. E., "An Empirical Model of Labor Supply in a Life-Cycle Setting," *The Journal of Political Economy* 89 (1981), 1059–85.

MAGNAC, T., AND D. THESMAR, "Identifying Dynamic Discrete Decision Processes," *Econometrica* 70 (2002), 801–16.

MANSO, G., "Experimentation and the Returns to Entrepreneurship," *The Review of Financial Studies* 29 (2016), 2319–40.

MARCH, J. G., AND Z. SHAPIRA, "Managerial Perspectives on Risk and Risk Taking," *Management Science* 33 (1987), 418–38.

MARGIOTTA, M. M., AND R. A. MILLER, "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review* 41 (2000), 418–38.

METRICK, A., "A Natural Experiment in 'Jeopardy!'," *American Economic Review* 85 (1995), 240–53.

MILLER, R. A., "Job Matching and Occupational Choice," *The Journal of Political Economy* 92 (1984), 1086–120.

MORRISEY, M. A., G. A. JENSEN, AND R. J. MORLOCK, "Small Employers and the Health Insurance Market," *Health Affairs* 13 (1994), 149–61.

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), 442–54.

RYAN, C., AND J. SIEBENS, "Educational Attainment in the United States: 2009," Technical Report, U.S. Census Bureau, 2012.

SCHUMPETER, J. A., *The Theory of Economic Development* (Cambridge, MA: Harvard University Press, 1911).

SHANE, S., "Why Encouraging More People to Become Entrepreneurs is Bad Public Policy," *Small Business Economics* 33 (2009), 141–9.

SKIRA, M. M., "Dynamic Wage and Employment Effects of Elder Parent Care," *International Economic Review* 56 (2015), 63–93.

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), 566–91.

VON GRAEVENITZ, G., D. HARHOFF, AND R. WEBER, "The Effects of Entrepreneurship Education," *Journal of Economic Behavior & Organization* 76 (2010), 90–112.