Online Appendix

The Macroeconomic Implications of Uncertainty and Learning for Entrepreneurship

Han Gao
UNSW

Lichen Zhang
HKU

A Data and Measurement

A.1 PSED

We use PSED-I (1998-2004) where there are 590 Nascent Entrepreneurs (NE) and 227 people in Controlled Group (CG) are surveyed. Variables related to businesses include business status, capital structure, legal form, expectations, and performances (sales/employment).

To be considered as a NE, individuals need to satisfy the following four criteria. First, the individual had to currently consider himself or herself as involved in the firm creation process. Second, he or she had to have engaged in some business startup activity in the past 12 months. Third, the individual had to expect to own all or part of the new firm being created. Fourth, the initiative, at the time of the initial screening survey, could not have progressed to the point that it could have been considered an operating business.

Key features of NEs in PSED In terms of legal forms, more than 84% are passthroughs. 50% of NE go with Sole Proprietorships, 20% go with Partnerships, 14% go with S-corp or LLC, 11% go with C-corp, 5% undecided. Regarding whether NEs are attached to paid jobs, about half of them have a paid job (partime or fulltime). 31% of men and 25% of women work full time on their new businesses (>= 35 hrs per week). Large majority of both sexes work for a paid job: Of the 70% of men working for pay, 55% did so full time. The analogous statistics for women are 62% and 39%. In terms of business size operated by NEs, around 40% of men and 50% of women choose to be "merely" self-employed, while the rest expect to become employers over the first five years of operation. As for the industrial choice, a large fraction of the men (35%) is starting a business in Health, Education, and Social services. Among the female NE this is also a strong category (20%). Retail and Restaurants account for 28% of the men and 45% of the women. 15% of the women and 8% of the men chose manufacturing.

Expectation formation and learning The learning process is captured by how forecast revision on a business' performance depends on the corresponding forecast error. We rely on the following questions from PSED I to measure forecast errors and forecast revision. Respondents in Wave 1 of PSED I report (1) *We would like to ask about your expectations regarding the future of this new firm. First, what would you expect the total sales, revenues, or fees to be in the first full year of operation?* (2)*And what about in the FIFTH year?*. Respondents in Wave 2-4 report (1) *What sales or revenue do you expect in the (current financial year/first full year of operation)?* [1] (2) *What annual sales or income would you expect for the firm FIVE years after the first full year of sales?* (3) *What annual sales or income would you expect for the firm TEN years after the first full year of sales?* This means we have (1) forecast data on sales for period 1 and 5 in period 0, (2) realized sales in period 1, 2, 3, and (3) forecast data on sales for period 5 and 10 in period 1,2,3.

(a) Forecasts (b) Realizations

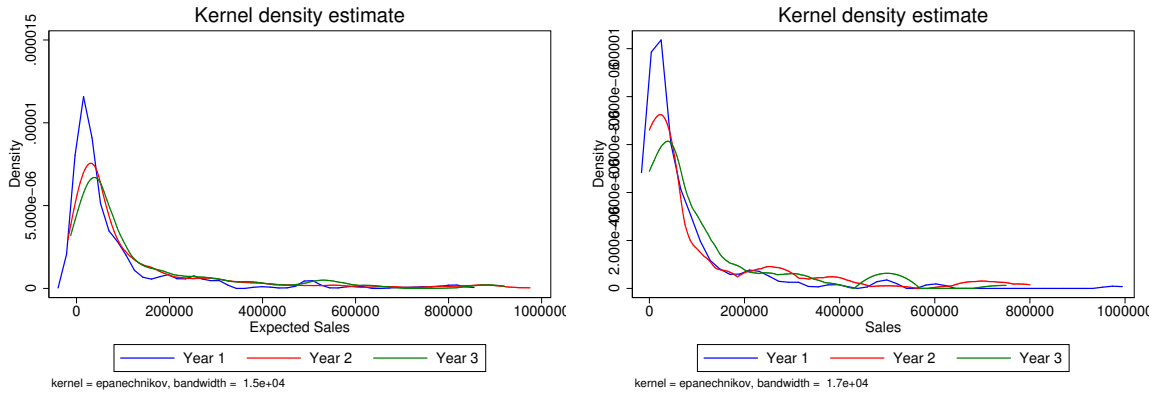

Figure A1: Expected and Realized Sales by Wave

(a) Sales (b) Employment (c) Correlation

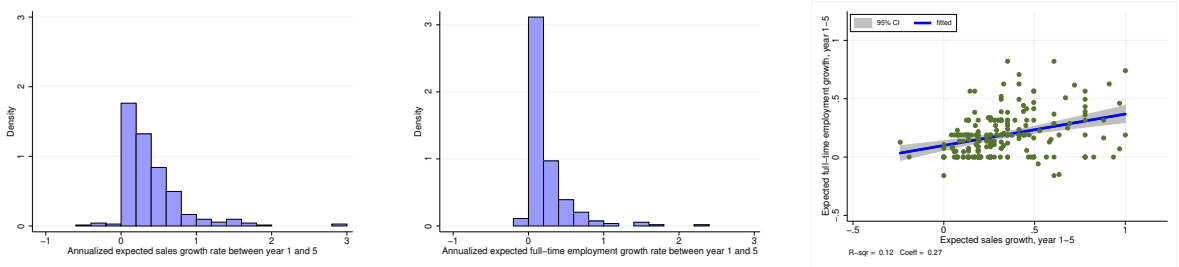

Figure A2: Distribution of Forecast Variables in PSED

[1] Since this question asks about sales/revenue in the current financial year, we regard the answer to this question as the realized sales in that year. To make this approximation solid, we only keep NE that have started operating businesses.

Figure A1 plot the forecasts and realizations of sales across three waves. As Figure A1 shows, both the distributions of forecasts and realized sales are quite dispersed, with long right tails. Besides sales, the PSED also asks forecasts of employment. Figure A2 plots the distribution of forecasts on sales and employment growth, as well as their correlation, respectively. As can be seen from Figure A2, sales and employment growth are positively correlated. However, the expected employment can only take integer values and we also observe a large fraction of entrepreneurs report a constant expected employment. In the data, approximately half of the NEs opt for "mere" self-employment, meaning they have no employees and no intention to expand their employment size. Therefore, we use sales as the primary object to measure entrepreneurs' belief and its evolution. In Table A1, we further presents key summary statistics of forecasts and realized sales.

Table A1: Summary Statistics of Sales in PSED

	Mean	25%	Median	75%	Max	Std. Dev.	Skewness	Frac. zero sales	Exit rate
Expected sales in wave 1 ($1000), conditional on entry									
Year 1	214	10	30	100	10,000	823	9.22	0.03	
Year 5	1,789	10	100	350	80,000	7,401	7.40	0.01	
Realized sales in following-up waves ($1000)									
Wave2	241	5	25	90	10,000	1,004	7.34	0.04	0.50
Wave3	508	10	25	185	25,000	2,817	8.38	0.03	0.16
Wave4	887	11	50	200	45,000	5,502	7.87	0.06	–

In Table A2, we compare the mean and standard deviation of forecast errors (FEs) by several key dimensions including gender, age, education, industry, and entrepreneur's past experience in terms of previous entrepreneurial activities, experience in the current industry, and experience in managerial occupations. We find that on average, female entrepreneurs and entrepreneurs with experience in the current industry have lower FE dispersion, while the difference in terms of FE dispersion is not significant in all other dimensions.

Figure 3(b) in the main text shows that FEs predict forecast revisions (FRs), captured by a simple linear regression with a slope of 0.67 for the whole sample. In Table A3, we extend the analysis by adding more control variables to the regression to see how the *learning speed* may potentially differ by entrepreneurs' characteristics. To be more specific, we add interactions of FEs with gender, education, retail industry, age polynomials, and the respondents' previous entrepreneurial experience as well as experience in the current industry. We find that operating in the retail industry is the only variable that significantly affects the learning speed, while all other variables have little impact, which rationalizes

Table A2: Distributions of Forecasting Errors by Characteristics of Entrepreneurs

Variable		Frac. obs.	Mean	Std. Dev.	Frac. missing	t-test	sd-test
Overall			-0.02	0.48			
Female	Yes	0.45	-0.03	0.44	0.01	0.76	0.13
	No	0.55	-0.01	0.52			
Age>=40	Yes	0.55	-0.04	0.50	0.05	0.54	0.58
	No	0.45	0.01	0.47			
College edu.	Yes	0.46	0.00	0.47	0.01	0.60	0.57
	No	0.54	-0.04	0.50			
Retail ind.	Yes	0.23	-0.03	0.48	0.00	0.90	0.99
	No	0.77	-0.02	0.49			
First business	Yes	0.40	0.01	0.43	0.50	0.97	0.66
	No	0.60	0.01	0.46			
Ind. exp.	Yes	0.86	0.03	0.45	0.50	0.50	0.70
	No	0.14	-0.07	0.40			
Manage exp.	Yes	0.18	-0.08	0.46	0.51	0.42	0.85
	No	0.82	0.03	0.45			

our choice of abstracting from detailed characteristics for both the empirics and the model in our main text. In particular, the fact that the initial distribution of forecast errors changes little with respect to entrepreneurs' age indicates that these errors likely pertain to uncertainty about entrepreneurial ability rather than other factors such as demand. This evidence further supports our "no cross-learning between occupations" assumption discussed in Section 2 of the main text.

Summarize survey questions on personality traits We use Principal Component Analysis to summarize the original 25 questions into several key traits. The construction follows Lise and Postel-Vinay (2020), which summarize multiple questions on detailed skills into three main skills. Consider number n types of main traits, the construction method is as follows:

1. Run PCA on PSED questions and keep the first n principal components;
2. Recover traits indices by recombining predicted principal components in such a way that they satisfy n certain exclusion restrictions;
3. Rescale the constructed traits to lie in [0,1].[2]

[2]Technical details are referred to the orginal paper of Lise and Postel-Vinay (2020).

Table A3: FR to FE Regression by Controls

	(1)	(2)	(3)	(4)
FE	0.606***	0.576*	0.642***	0.724*
	(5.19)	(1.89)	(3.01)	(1.88)
FE× female	0.00531	0.0215	0.162	0.353
	(0.04)	(0.14)	(0.67)	(1.37)
FE× college	0.00546	-0.00310	-0.243	-0.222
	(0.04)	(-0.02)	(-1.01)	(-0.90)
FE× retail	0.293*	0.310*	0.359	0.183
	(1.75)	(1.76)	(0.97)	(0.49)
FE× age		0.000339		
		(0.05)		
FE× age^2		-2.86e-08		
		(-0.04)		
FE× first business			-0.308	
			(-1.23)	
FE× log exp				-0.109
				(-0.86)
Personality controls	No	Yes	No	No
Obs.	146	146	72	61
adj. R^2	0.395	0.373	0.254	0.269

Note: t-statistics are reported in parentheses.

Besides the five traits considered by Hamilton, Papageorge, and Pande (2019), we additionally consider a general trait for running a business.[3] This is to isolate the preference solely for doing business, which is orthogonal to the general OCEAN traits such as risk-taking, social activities, etc.. The restrictions we consider are the questions in column 'Restriction' of Table A4 only reflect the corresponding traits.

Correlations between traits Table A5 reports the correlation between the constructed personality traits.

Table A6 describes differences in the constructed personality traits between men and women and between different age groups. Men have significantly higher scores in openness traits and lower scores in extraversion. Older individuals have significantly higher scores in conscientious trait. All other comparisons are not statistically significant. In particular, the love of business trait does not differ by gender or age.

Note that we are not the first to use the data from PSED to document empirical facts

[3]More details on "OCEAN" can be found on https://en.wikipedia.org/wiki/Big_Five_personality_traits.

Table A4: Correspondence between "OCEAN" and Survey Questions

Personality traits	Description	Restriction
Love of business	general love of business	QL1d
Openness to experience	inventive/curious vs. consistent/cautious	QL1q
Conscientiousness	efficient/organized vs. extravagant/careless	QL1b
Extraversion	outgoing/energetic vs. solitary/reserved	QL1h
Agreeableness	friendly/compassionate vs. critical/rational	QL1x
Neuroticism	sensitive/nervous vs. resilient/confident	QL1i

Table A5: Correlations between Personality Traits

	LoB	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Love of Business	1.0000					
Openness	0.3606	1.0000				
Conscientiousness	0.3237	0.3368	1.0000			
Extraversion	0.1182	0.0056	0.3695	1.0000		
Agreeableness	0.3206	0.5560	0.2088	0.6670	1.0000	
Neuroticism	0.2347	0.6973	0.6665	0.6044	0.8184	1.0000

related to non-pecuniary benefits that determines the entry of entrepreneurship. Hurst and Pugsley (2011) also use PSED and show that the median small business reports starting their business for non-pecuniary reasons. However, their approach is different from ours. They rely on the question *"Why do [or did] you want to start this new business?"*. They took the raw responses to the question and created five broad categories of their own including non-pecuniary reasons and reasons related to the generation of income. The main responses in the non-pecuniary category include "want to be my own boss," "flexibility/set own hours," "work from home," and "enjoy work, have passion for it/ hobby." They find that roughly 50 percent of all respondents reported non-pecuniary benefits as being one of the primary reasons they started their business. We see the results generated using our approach complementary to theirs, and the biggest advantage of our approach is that we can generate a distribution of Love of Business characteristic, which can be used to discipline the non-pecuniary utility in our structural model.

Personality traits and entrepreneurship We plot the distribution of scores of the six personal traits for two groups of individuals in our sample—nascent entrepreneurs and workers (control group). As shown in Figure A3, only the distribution of Love of Business

Table A6: Comparison of Personality Traits by Gender and Age

	By gender			By age		
	Men	Women	p-value	Age < 40	Age ≥ 40	p-value
Love of Business	0.5742	0.5749	0.9538	0.5727	0.5774	0.7189
	(0.0086)	(0.0085)		(0.0088)	(0.0094)	
Openness	0.5016	0.4685	0.0018	0.4823	0.4871	0.6694
	(0.0078)	(0.0072)		(0.0075)	(0.0083)	
Conscientiousness	0.6021	0.6237	0.0410	0.6250	0.6006	0.0311
	(0.0074)	(0.0075)		(0.0083)	(0.0076)	
Extraversion	0.5623	0.6117	0.0000	0.5847	0.5876	0.7984
	(0.0071)	(0.0072)		(0.0078)	(0.0079)	
Agreeableness	0.6203	0.6237	0.7123	0.6174	0.6270	0.3297
	(0.0065)	(0.0066)		(0.0067)	(0.0072)	
Neuroticism	0.5912	0.5946	0.7235	0.5945	0.5908	0.7106
	(0.0063)	(0.0067)		(0.0069)	(0.0071)	
Sample size	379	395		337	337	

scores exhibit significant difference between the two groups of people.

We run Heckman Two-Step Regression to further explore how LoB affects the entrepreneurial choice. In the first stage, we run the following regression:

$$Probit\left(E = 1|Z\right) = \Phi\left(Z\gamma\right)$$

where $E = 1$ if the respondent is an entrepreneur and $E = 0$ otherwise. In the second stage, we run the following regression:

$$entrep.\, income = X\beta + u$$

We report the results in Table A7.

A.2 PSID

The PSID sample used for studying the life-cycle behavior of entrepreneurs was generated following Heathcote, Perri, and Violante (2010) in general. From the raw data, we extract a sample of heads of households from the SRC sample based on the waves from 1970 to 1997.

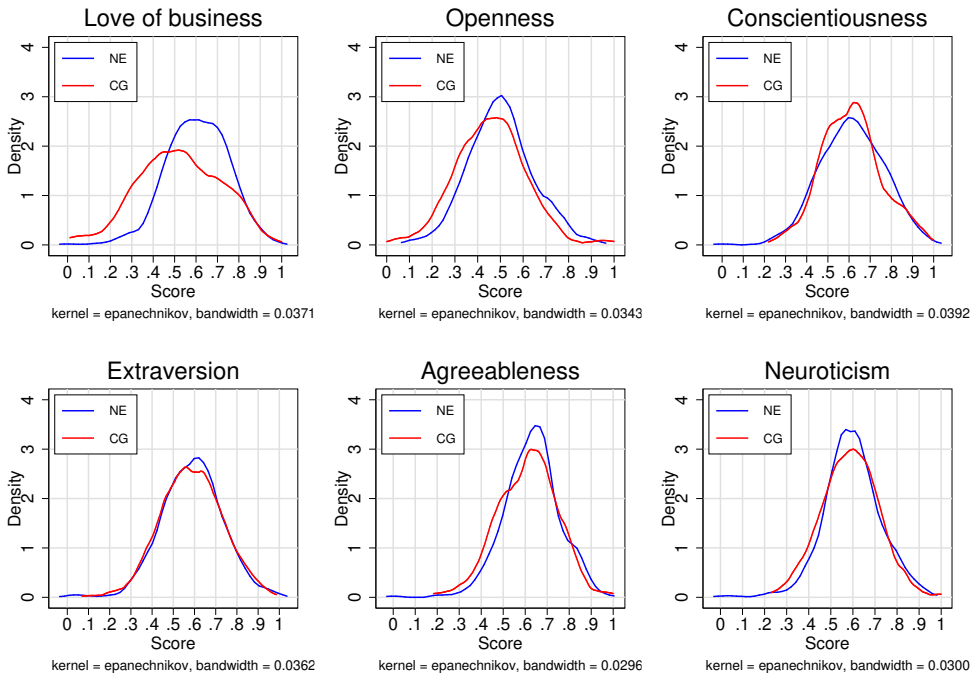

Figure A3: Distribution of Personal Traits Scores

All monetary variables (income and wealth) are deflated using the Personal Consumption Expenditure index (PCE) and expressed in 2010 dollars. The baseline sample considers households whose head is between 21 and 65 years old, both ends included. We report summary statistics of the sample in Table A8.

Definition of "head" The head of the family unit (FU) must be at least 16 years old, and the person with the most financial responsibility in the FU. If this person is female and she has a husband in the FU, then he is designated as head. If she has a boyfriend with whom she has been living for at least one year, then he is head. However, if she has 1) a husband or a boyfriend who is incapacitated and unable to fulfill the functions of head, 2) a boyfriend who has been living in the FU for less than a year, 3) no husband/boyfriend, then the FU will have a female head. A new head is selected if last year's head moved out of the household unit, died or became incapacitated, or if a single female head has gotten married. Also, if the family is a split-off family (hence a new family unit in the sample), then a new head is chosen.

Samples In this paper, we only consider SRC sample (i.e. id68 <= 3000).

Table A7: Heckman Two-Step Regression

		OLS		Heckman Two-step	
	Love of Business	0.24	- 0.27	0.21	- 0.49
	Openness	- 1.00	- 0.81	- 0.99	- 0.71
	Conscientiousness	- 0.43	- 1.21	- 0.38	- 0.78
	Extraversion	- 1.68	- 0.78	- 1.70	- 0.76
Heckman Stage 2	Agreeableness	- 0.43	- 2.13	- 0.35	- 1.61
	Neuroticism	2.90	4.50	2.80	3.64
/OLS	Age/100	- 0.11	- 0.11	- 0.11	- 0.11
	log(experience)		0.11		0.11
	College	-0.02	- 0.13	- 0.02	- 0.15
	Female	0.05	- 0.13	0.05	- 0.15
	White	0.25**	0.24	0.24	0.21
	Love of Business			2.97***	3.42***
	Openness			- 0.93	- 1.64
	Conscientiousness			- 4.91	- 6.90
	Extraversion			1.24	- 0.32
	Agreeableness			- 7.89	- 8.54
Heckman Stage 1	Neuroticism			10.20	13.95
	Age/100			- 0.02	0.33
	log(experience)				- 0.15**
	White			0.36***	0.52***
	College			0.27**	0.30**
	Female			0.08	0.30*
	Observations	141	70	773	540
	R^2	0.1143	0.1251		

Notes: *,**,*** refer to significance at 10%, 5%, and 1% respectively.

Top-coding and bracketed variables We deal with top-coded observations by assuming the underlying distribution for each component of income is Pareto, and by forecasting the mean value for top-coded observations by extrapolating a Pareto density fitted to the non-top-coded upper end of the observed distribution.

In some of the early waves, a number of income measures were bracketed. For these variables, we use the midpoint of each bracket, and 1.5× the top-coded thresholds for observations in the top bracket.

Variable definitions In the PSID all the questions are retrospective, i.e., variables in survey—year t refer to calendar year $t-1$. The interview is usually conducted around March. When variables were not defined consistently across years (for example employment status was categorized differently in different years), the variables were recoded based on

Table A8: Summary Statistics of PSID sample

	Wage Workers	Entrepreneurs	Labor Force	Total
Obs. per year	2,284	261	2,690	2,994
Age (mean)	38.1	43.0	38.4	39.6
Men (%)	83.4	94.6	83.7	81.4
College or above (%)	27.2	36.1	27.3	26.0
White (%)	89.7	96.0	89.8	89.1
Income (mean, 2010$)	49,357	74,778	50,135	45,882
Wealth (mean, 2010$)	153,164	688,013	206,887	206,294

Notes: The table reports statistics of a sample of heads of households between 21 and 65 years old. Each statistic is the sample average across all the survey waves between 1970 and 1997. Entrepreneurs are defined as self-employed business owners. All monetary values are deflated by the PCE index and expressed in 2010 US dollars.

their original (and less detailed) coding, so as to be consistent across years.

Income and earnings: Labor income of heads is defined as income from wages, salaries, commissions, bonuses, overtime and the labor part of self-employment income. The PSID splits self-employment income into asset and labor components using a 50-50 rule.

The earnings of heads consists of both labor income and business income, which is equal to the labor income of head plus the asset part of business income. Note that the variable on the asset part of business income only applies to individuals who runs unincorporated businesses. Unincorporated business owners are not sheltered from the losses of their ventures through limited liability. This means that a head's income can be positive, zero, or negative.

Wealth: The measure of wealth is the variable WEALTH2, which is available in specific waves of PSID. This variable is constructed as sum of values of several asset types (family farm business, family accounts, assets, stocks, houses, and other real estate etc.) net of debt value.

Annual hours of work is defined as the sum of annual hours worked on the main job, on extra jobs, plus annual hours of overtime. It is computed by the PSID using information on usual hours worked per week and the number of actual weeks worked in the last year.

Labor force: a household head is considered in the labor force if her employment status is either "Working now", "Only temporarily laid off, sick leave or maternity leave", or "Looking for work, unemployed".

Entrepreneur: The PSID provides several questions that can be used to classify individuals' entrepreneurial status. In our analysis, we use two of these questions. The first question is "Did you (or anyone else in the family there) own a business at any time in (year) or have a financial interest in any business enterprise?". The second one is "On your main job, are you (head) self-employed, or are you employed by someone else?". An individual is defined as an entrepreneur if her answer to both questions are "yes".

Worker: a household head is considered to be worker if (1) her employment status is "Working now" or "Only temporarily laid off, sick leave or maternity leave", (2) she is neither self-employed nor a business owner, (3) her labor income is positive, and (4) her annual hours is greater than 260.

Retirement: a household head is considered to be retired if (1) her employment status is "Retired", and (2) her social security income is positive. Note that adding condition (2) is to avoid the misreport of retirement status. If we only rely on condition (1) to define retirement, we will see a pattern that around 5% of retirees in our sample are within the age group of 21-50.

Here is the detailed procedure about constructing variables household earnings and hourly wages:

1. Obtain the SRC sample that includes data for labor income, business income, employment status, gender, age, education, race, wealth, indicator on business owner for heads and wives of households.

2. Drop any observation (household) with missing age for either head or spouse.

3. Drop any observation with missing earnings but positive annual hours of work.

4. Drop any observation with positive earnings but zero annual hours of work.

5. Drop any observation with either head or spouse has nominal wage below half of the minimum wage.

6. Drop any household if neither the head nor the spouse is of working age, which we define as between the ages of 21 and 65.

A.2.1 Earnings and wealth from PSID

In this section, we plot the earnings and wealth over the life-cycle as well as the earnings distributions for different groups of people. We consider three groups: (1) entrants, which means first-year entrepreneurs, (2) incumbents which means entrepreneurs excluding entrants, and (3) workers. Earnings is defined as above. That is, workers' earnings are

their labor income. Entrepreneurs' earnings are their labor income plus business income. The measure of wealth is the variable WEALTH2 as found in specific waves of PSID. This variable is constructed as sum of values of several asset types (family farm business, family accounts, assets, stocks, houses, and other real estate etc.) net of debt value.

In Figure A4(a), we plot the median earnings for entrants, incumbents, and workers over the life-cycle respectively, with the median earnings of workers with age 26-30 normalized to one. We can see that the median earnings of entrant entrepreneurs is always smaller than that of workers over the life-cycle. This may suggest non-pecuniary value of entry, which is consistent with the two elements—learning and Love of Business characteristics—in our model.[4]

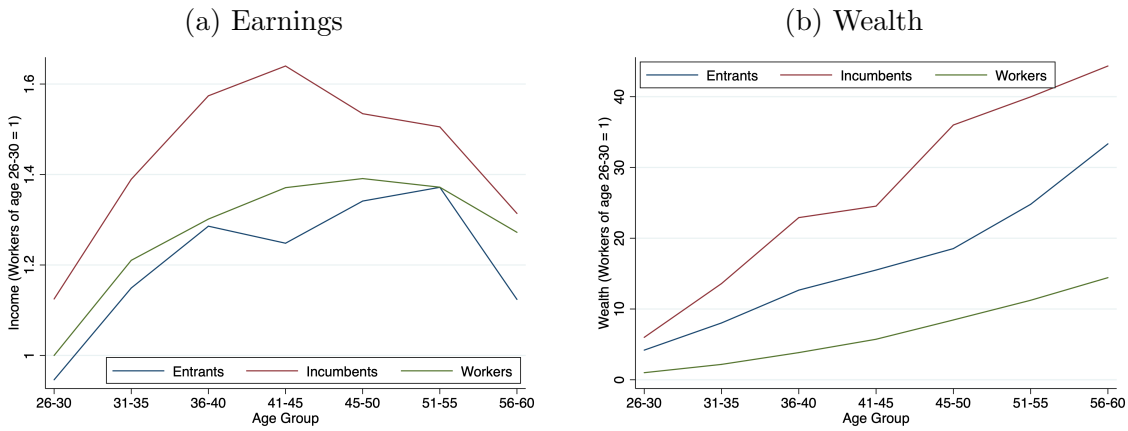

Figure A4: Median Earnings and Wealth over the Life Cycle

In Figure A4(b), we plot the median wealth for entrants, incumbents, and workers over the life-cycle respectively, with the median wealth of workers with age 26-30 normalized to one. It is not surprising to see that entrepreneurs have higher median wealth level compared to workers, which is consistent with the stylized facts that entrepreneurs in general are relatively wealthier people. For example, in the SCF, even though households headed by entrepreneurs make up only 7 to 8 percent of the population, they own nearly one-third of the wealth in the United States.

In Figure A5, we plot the distributions of earnings for entrants, incumbents, and workers respectively. We can see that both the earnings distributions of entrants and incumbents are more dispersed than that of workers, with the median of entrants' earnings smaller.

[4]We also admit that this pattern may be due to some mechanical reasons. For example, if a person enters entrepreneurship in October and is considered as an entrant entrepreneur in that year, then her earnings equals ten months' worker income and two months' entrepreneurial income. However, due to the limitations of PSID, we cannot rule out this kind of possibility.

While median earnings in entrepreneurs are lower than median wage earnings, a subset of entrepreneurs have very high earnings. This may suggest a learning story, as in, for instance, Hincapié (2020) and Dillon and Stanton (2018), that workers seek to maximize expected lifetime earnings may rationally enter entrepreneurship to learn about their entrepreneurial ability, with the option to exit entrepreneurship as uncertainty resolves, even though their realized earnings during entrepreneurship are often low.

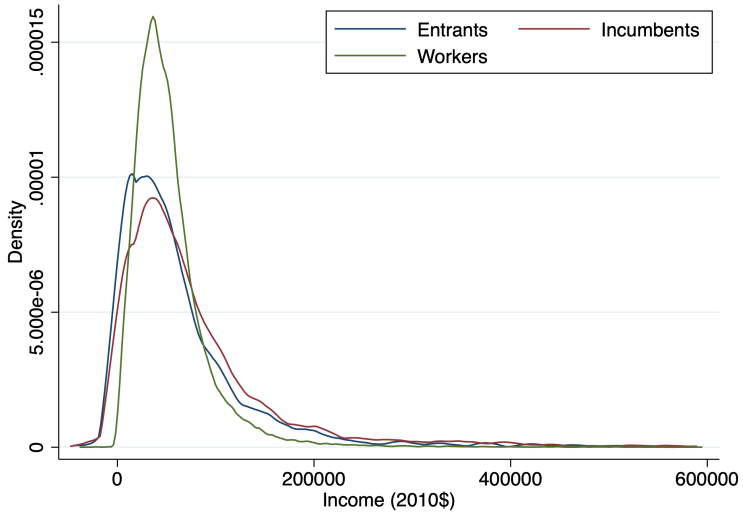

Figure A5: Earnings Distribution in PSID

A.2.2 Life-cycle patterns of entrepreneurship from the CPS

In order to verify the robustness of the life cycle patterns of entrepreneurship, we consider the Current Population Survey (CPS) that covers a much larger number of households compared to PSID. We construct a panel using monthly CPS data following the method developed by Drew, Flood, and Warren (2014). From the raw data, we extract a sample of heads of households from 1976 to 1997 at a monthly basis. All monetary variables (income and wealth) are deflated using the Personal Consumption Expenditure index (PCE) and expressed in 2010 dollars. The baseline sample considers households whose head is between 21 and 65 years old, both ends included. We report summary statistics of the sample in Table A9.

The entry and exit rate can thus only be computed at monthly frequency. We try to make our CPS sample as close to our benchmark PSID sample as possible. The CPS sample covers a similar periods (1975 - 1997) and the entrepreneurs in CPS are defined as self-employed household heads.[5] The age-profiles of entry, exit, and entrepreneurs share

[5]There is no variable on whether an individual is a business owner or not in CPS.

Table A9: Summary Statistics of CPS Sample

	Wage Workers	Entrepreneurs	Labor Force	Total
Obs. per month	32,018	4,777	36,137	42,504
Age (mean)	40.1	44.4	40.5	41.9
Men (%)	75.0	89.9	77.1	73.3
College or above (%)	25.2	29.2	26.0	23.7

Notes: The table reports statistics of a sample of heads of households between 21 and 65 years old. Each statistic is the sample average across all the survey waves between 1976 and 1997 at a monthly basis. Entrepreneurs are defined as self-employed heads of households. All monetary values are deflated by the PCE index and expressed in 2010 US dollars.

using CPS are reported in Figure A6. Although the numbers are not directly comparable between figures using the two datasets due to different definitions of entrepreneurs and different data frequencies, their life-cycle patterns are extremely similar to each other.

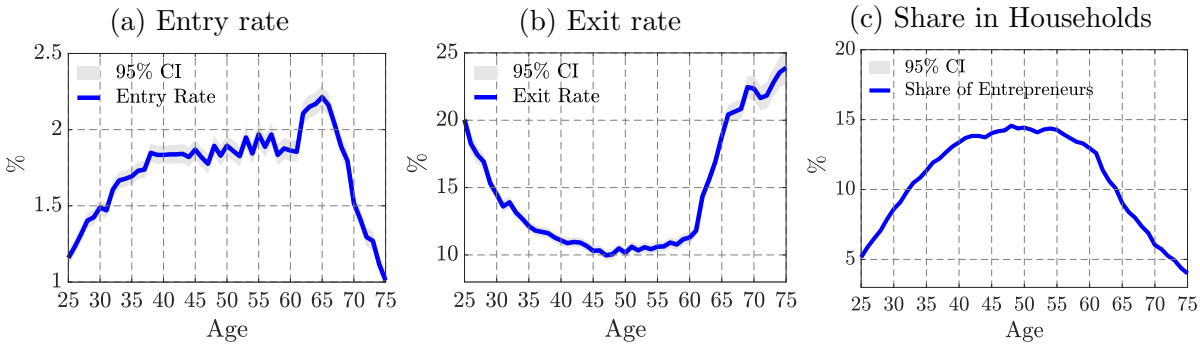

Figure A6: Entrepreneurship over the Life Cycle, CPS Sample

A.3 SCF

We consider two kinds of definitions of business income in SCF and check the share of negative or non-positive business income over the life cycle. The results are reported in Figure A7. In Definition 1, business income = schedule-C business income + taxable interest income + dividend income + capital gains + schedule-E business income + net operating loss. In Definition 2, business income = schedule-C business income + schedule-E business income.

We also document age profiles of the entrepreneur share, earnings, and wealth in SCF, as in Figure A8. The aggregate entrepreneur share over the life cycle is consistent with the patterns in PSID and CPS where we use the same definition of entrepreneurs as the PSID.

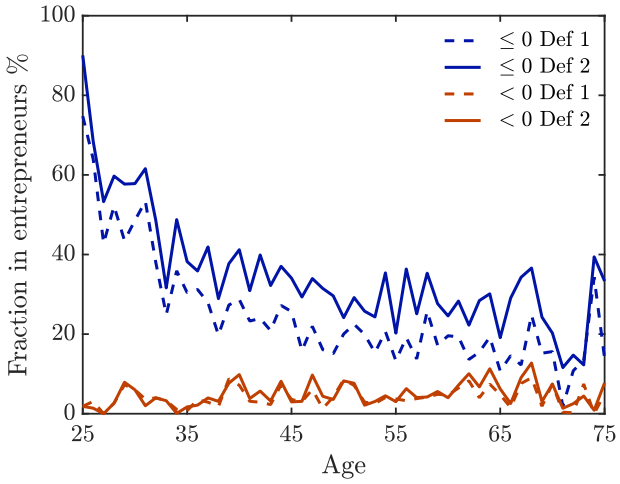

Figure A7: Non-positive Business Incomes in SCF

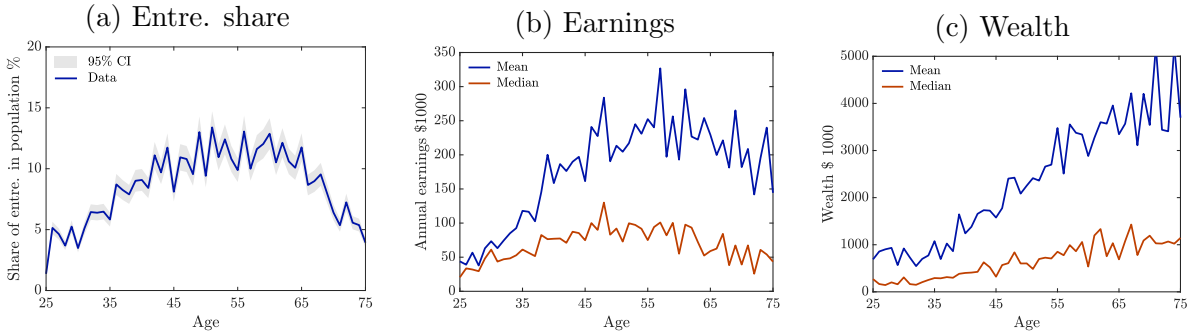

Figure A8: Age Profile of Entrepreneurial Share, Incomes, and Assets in SCF

Finally, we compare several key statistics across PSID, SCF, and PSED as in Table A10. Notes: In IRS integrated business data, share of unincorporated is around 79% in 1996. Among all corporations, around 50% are s-corps.

Table A10: Comparison of Entrepreneurs Sample across PSID, SCF, and PSED

	PSID (96-04)	SCF (97-03)	PSED (98-04)
Frac. of entrep. who have wage income	60%	77%	66%
Frac. of entrep. whose businc>0.5*total inc	49%	56%	-
Share of unincorporated	67%	75%	>70%
Exit rate after 1 year operation	29%	-	30%

B Model

B.1 Details of the Recursive Problems

Value in Normal Working Age $(0 < j < J^V)$ During normal working ages, individuals make occupational choice decisions between being a worker or entrepreneur. For $o = W$,

$$
\begin{aligned}
V_j^W(x_e, \chi_w, a, \epsilon_w, \tilde{\mu}_e, \tilde{\nu}_e, \epsilon_e) = \max_{a',c,l} \{ & u(c, l; x_e) \\
& + \beta[(1-\psi_j) \max_{o' \in \{W,E\}} \{\mathbb{E}V_{j+1}^W(x_e, \chi_w, a', \epsilon'_w, \tilde{\mu}'_e, \tilde{\nu}'_e, \epsilon'_e), \mathbb{E}V_{j+1}^E(x_e, \chi_w, a', \epsilon'_w, \tilde{\mu}'_e, \tilde{\nu}'_e, \epsilon'_e)\} + \psi_j \mathcal{V}(a')]\} \\
s.t. \quad & a' + c(1+\tau_c) = a(1+r) + (1-\tau_{ss}) y_j^o(a, \epsilon_w, \epsilon_e) - T^o(y_j^o + ra) - \kappa_e \mathbb{1}_{\{o'=E\}} \\
& \{\tilde{\mu}'_e, \tilde{\nu}'_e\} = \{\tilde{\mu}_e, \tilde{\nu}_e\} \\
& a' \geq \underline{a},
\end{aligned}
\tag{PA1}
$$

where $y_j^o(a, \epsilon_w, \epsilon_e)$ is the total o-occupation pre-government income, κ_e is the fixed entry cost, and $\mathbb{1}_{\{o'=E\}}$ is the corresponding indicator function that specifies if households who switch occupations to entrepreneurs next period need to pay such a cost.

For $o = E$,

$$
\begin{aligned}
V_j^E(x_e, \chi_w, a, \epsilon_w, \tilde{\mu}_e, \tilde{\nu}_e, \epsilon_e) = \max_{a',c,l} \{ & u(c, l; x_e) \\
& + \beta\delta_e[(1-\psi_j)\mathbb{E}V_{j+1}^W(x_e, \chi_w, a', \epsilon'_w, \tilde{\mu}'_e, \tilde{\nu}'_e, \epsilon'_e) + \psi_j \mathcal{V}(a')] \\
& + \beta(1-\delta_e)[(1-\psi_j) \max_{o' \in \{W,E\}} \{\mathbb{E}V_{j+1}^W(x_e, \chi_w, a', \epsilon'_w, \tilde{\mu}'_e, \tilde{\nu}'_e, \epsilon'_e), \mathbb{E}V_{j+1}^E(x_e, \chi_w, a', \epsilon'_w, \tilde{\mu}'_e, \tilde{\nu}'_e, \epsilon'_e)\} + \psi_j \mathcal{V}(a')] \\
s.t. \quad & a' + c(1+\tau_c) = a(1+r) + (1-\tau_{ss}) y_j^o(a, \epsilon_w, \epsilon_e) - T^o(y_j^o + ra) \\
& \{\tilde{\mu}'_e, \tilde{\nu}'_e\} = \Pi(\tilde{\mu}'_e, \tilde{\nu}'_e | \tilde{\mu}_e, \tilde{\nu}_e, \epsilon_e) \\
& a' \geq \underline{a},
\end{aligned}
\tag{PA2}
$$

where δ_e is the exogenous separation shock that only applies to current incumbent entrepreneurs.

B.2 Definition of the Stationary Competitive Equilibrium

An individual with age j is indexed by states $\mathbf{x}_j = (x_e, \chi_w,\ a_j,\ \epsilon_{w,j},\ \tilde{\mu}_{e,j},\ \tilde{\nu}_{e,j},\ \epsilon_{e,j})$. Given a tax structure $\left\{\tau_c, T^\omega(\cdot), T^b(\cdot), \tau_{ss}\right\}$ and an initial distributions of workers and entrepreneurs over individual states $\left\{\Gamma_0^W(\mathbf{x}_0), \Gamma_0^E(\mathbf{x}_0)\right\}$, a **stationary recursive competitive equilibrium** comprises

- prices $\{w, r\}$ and social security benefits z

- a sequence of workers' policy functions on saving, occupation choice, consumption, and hours, $\left\{a_W^{'}\left(\mathbf{x}_j\right), o_W^{'}\left(\mathbf{x}_j\right), c_W^{'}\left(\mathbf{x}_j\right), h\left(\mathbf{x}_j\right)\right\}_{j=1}^{J^R-1}$, with associated value functions $\left\{V_j^W\right\}_{j=1}^{J^R-1}$, a sequence of entrepreneurs' policy functions on saving, occupation choice, consumption, capital rental, and labor hired, $\left\{a_E^{'}\left(\mathbf{x}_j\right), o_E^{'}\left(\mathbf{x}_j\right), c_E^{'}\left(\mathbf{x}_j\right), k\left(\mathbf{x}_j\right), n\left(\mathbf{x}_j\right)\right\}_{j=1}^{J^R-1}$, with associated value functions $\left\{V_j^E\right\}_{j=1}^{J^R-1}$, and individuals' policy functions after retirement on saving and consumption, $\left\{a_R^{'}\left(\mathbf{x}_j\right), c_R^{'}\left(\mathbf{x}_j\right)\right\}_{j=J^R}^{J}$, with associated value functions $\left\{V_j^R\right\}_{j=J^R}^{J}$
- factors demand of the corporate sector, $\{K_C, N_C\}$
- a sequence of distributions over idiosyncratic states for both workers and entrepreneurs $\left\{\Gamma_j^W\left(\mathbf{x}_j\right), \Gamma_j^E\left(\mathbf{x}_j\right)\right\}_{j=1}^{J}$

such that

1. Given prices w, r, the tax structure $\left\{\tau_c, T^{\omega}(\cdot), T^b(\cdot), \tau_{ss}\right\}$, and social security benefits z, the policy functions solve individual's problems (P1) and (P2).
2. The factors demand of the corporate sector solve equation (8).
3. Capital market clears:

$$\sum_{j=1}^{J}\int a^W(\mathbf{x}_j)d\Gamma_j^W(\mathbf{x}_j)+\sum_{j=1}^{J}\int a^E(\mathbf{x}_j)d\Gamma_j^E(\mathbf{x}_j)=K_C+\sum_{j=1}^{J^R-1}\int k(\mathbf{x}_j)d\Gamma_j^E(\mathbf{x}_j) \quad \text{(A1)}$$

4. Labor market clears:

$$\sum_{j=1}^{J^R-1}\int \epsilon_{\omega,j}\theta_j h_j(\mathbf{x}_j)\mathbb{I}_{\{h_j>0\}}d\Gamma_j^W(\mathbf{x}_j)=N_C+\sum_{j=1}^{J^R-1}\int n(\mathbf{x}_j)d\Gamma_j^E(\mathbf{x}_j) \quad \text{(A2)}$$

5. The Social Security system clears:

$$\tau_{ss}\left(\sum_{j=1}^{J^R-1}\int y_j^{\omega}(\mathbf{x}_j)d\Gamma_j^W(\mathbf{x}_j)+\sum_{j=1}^{J^R-1}\int y_j^{b}(\mathbf{x}_j)d\Gamma_j^E(\mathbf{x}_j)\right)=\sum_{j=J^R}^{J}z \quad \text{(A3)}$$

6. The government balances its budget:

$$G=\tau_c C+\sum_{j=1}^{J^R-1}\int T^{\omega}\left(y_j^{\omega}(\mathbf{x}_j)\right)d\Gamma_j^W(\mathbf{x}_j)+\sum_{j=1}^{J^R-1}\int T^{b}\left(y_j^{b}(\mathbf{x}_j)\right)d\Gamma_j^E(\mathbf{x}_j) \quad \text{(A4)}$$

7. The distributions of workers and entrepreneurs at the beginning of period j respectively, $\left\{\Gamma_j^W\left(\mathbf{x}_j\right), \Gamma_j^E\left(\mathbf{x}_j\right)\right\}_{j=1}^J$, evolve based on the individuals' policy functions and the autoregressive process for the exogenous productivity states.

B.3 Additional Model Fits of the Benchmark Model

Entrepreneurial Earnings The existing literature that incorporates learning in a structural model of entrepreneurs (e.g., Hincapié (2020); Dillon and Stanton (2018)) typically relies on moments of earnings for entrepreneurs and workers to identify the learning process. We use relatively direct evidence on entrepreneurs' expectation formation to discipline the learning process. In Figure A9, we demonstrate that our model can well replicate the mean and standard deviation of entrepreneurial income with respect to the entrepreneurial spell.

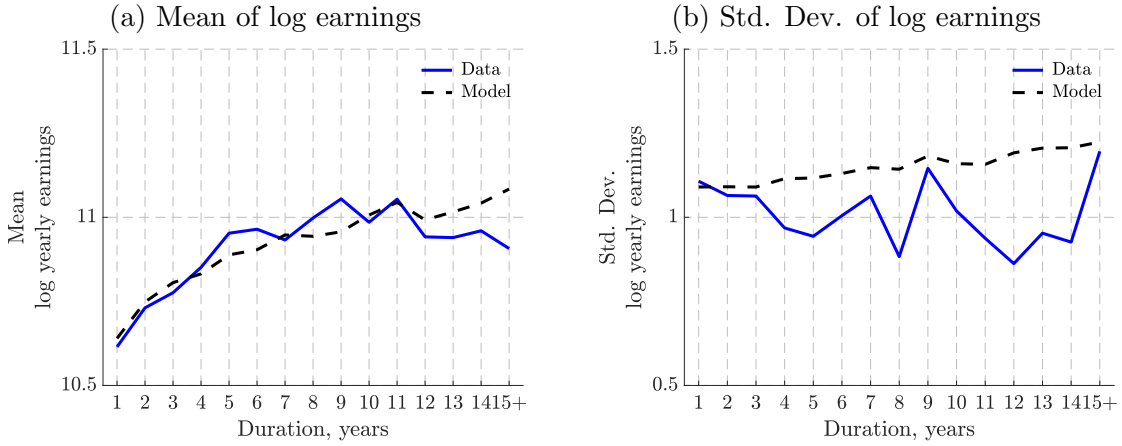

Figure A9: Model Fit: Earnings by Entrepreneurial Duration

Income and Wealth Distributions Our calibrated model captures both the income and wealth distribution well, as in Cagetti and De Nardi (2006). The results are reported in Table A11.

First time entry In Figure A10, we plot the first time entrepreneurs as a share of total population which contains overlapping information with the figure on the entry rate over the life cycle. In addition, we compare the moments implied by the benchmark model with the case of perfect information.

Aggregate moments We further check if the moments on the macroeconomic level generated from our model is consistent with the data. The results are reported in Table A12.

Table A11: Model Fit: Income and Wealth Distribution

	Benchmark	Data
Gini coefficient		
Income – all	0.52	0.55
Income – worker	0.35	0.38
Income – entrepreneur	0.62	0.66
Wealth – all	0.64	0.85
Income/wealth ratios: entrepreneur to worker		
Income – median	1.25	1.30
Income – mean	2.12	2.50
Wealth – median	5.90	6.00
Fraction of entrepreneurs in wealth percentiles		
Top 1%	0.56	0.54
Top 5%	0.48	0.39
Top 10%	0.31	0.32
Top 20%	0.22	0.22

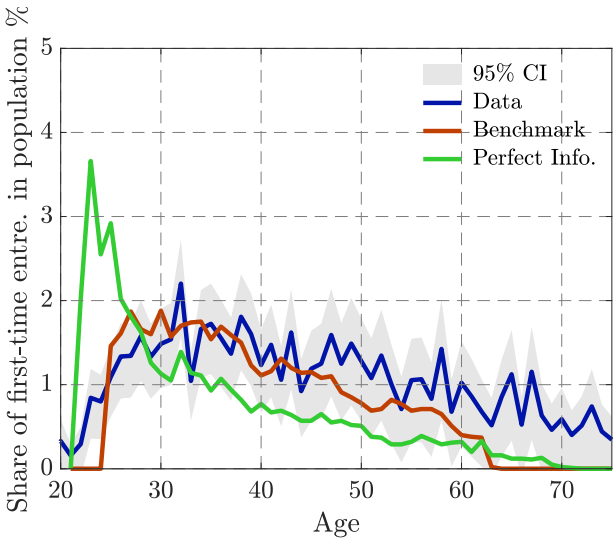

Figure A10: Model Fit: First-time Entrepreneurs as a Share of Total Population

Firm distribution in entrepreneurial sector We finally check the model fitness in distribution of firm size in terms of employment in the entrepreneurial sector. The statistics are reported in Table A13. We can see that our model is able to reproduce similar patterns to the empirical results we obtain from the Survey of Consumer Finance (SCF).

Table A12: Aggregate Moments

	Values
Taxes to GDP ratios, %	
Total taxes	23.9
Consumption tax	2.4
Wage income tax	16.6
Business income tax	1.6
Assets/sales to GDP ratios, %	
Corporate fixed asset	261.6
Entrepreneurial fixed assets	48.3
Entrepreneurial sales	21.3

Table A13: Model Fit: Firm Size Distribution of Entrepreneurs

	Data	Model
Share of entre. in population %	8.8	8.4
Share of hiring entre. %	66.1	82.9
Firm size distribution %		
1-5 Employees	69.2	42.3
6-10 Employees	11.9	40.2
11-20 Employees	6.5	17.5
>20 Employees	12.5	0.0

Exit of entrepreneurs around retirement For the increase in the exit rate after the age of 60, we do a decomposition of the exit rate to distinguish between the exit due to retirement and the exit due to switching occupation. The results are presented in Figure A11. Starting from the age of 62, increasing exit rate is only driven by the increasing retirement of people. Our model well replicates the exit of entrepreneurs around the retirement. The key element that helps to match the data is voluntary retirement and bequest.

Wealth percentiles As shown in Figure A12 using the Survey of Consumer Finance, there is still high level of wealth accumulation at the later stage of individuals' life cycle. By incorporating the element of bequest into our model, the wealth percentiles over the life cycle for all the individuals and for entrepreneurs only are close to its empirical counterpart.

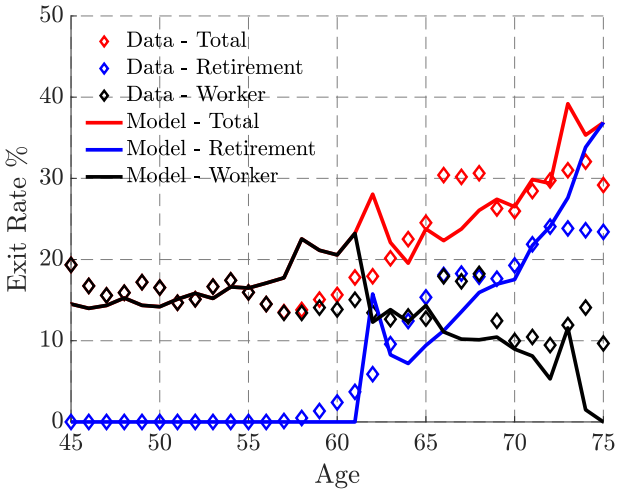

Figure A11: Model Fit: Exit of Entrepreneurs around Retirement

Panel (a) is the data and panel (b) is generated from the model. Our model can replicate the overall life cycle pattern of asset accumulation. To be more specific, in both the data and our model, for the overall population, asset peaks for the age group of 60-64 and drops afterwards. Our model slightly overpredicts the drop in asset for the entrepreneurs at older ages. Our model does a good job in matching the wealth distribution for both the overall population and the entrepreneurs. Our model slightly underestimates the gaps between 95% percentile and the median.

Dispersion of LoB characteristic Results are shown in Table A14.

Table A14: Love of Business Characteristic by Entrepreneur Status

	All	Workers		Entrepreneurs	
		Data	Model	Data	Model
Mean	0.531	0.521	0.524	0.614	0.612
Std. Dev.	0.190	0.193	0.189	0.123	0.171

B.4 Implications of Benchmark vs Perfect Information Model

B.4.1 Re-estimation under the Case of Perfect Information

In this section, we re-calibrate a version of our model, which we refer to as the perfect information case. The only deviation from the benchmark in the perfect information case is that individuals are already aware of their innate entrepreneurial ability upon entering

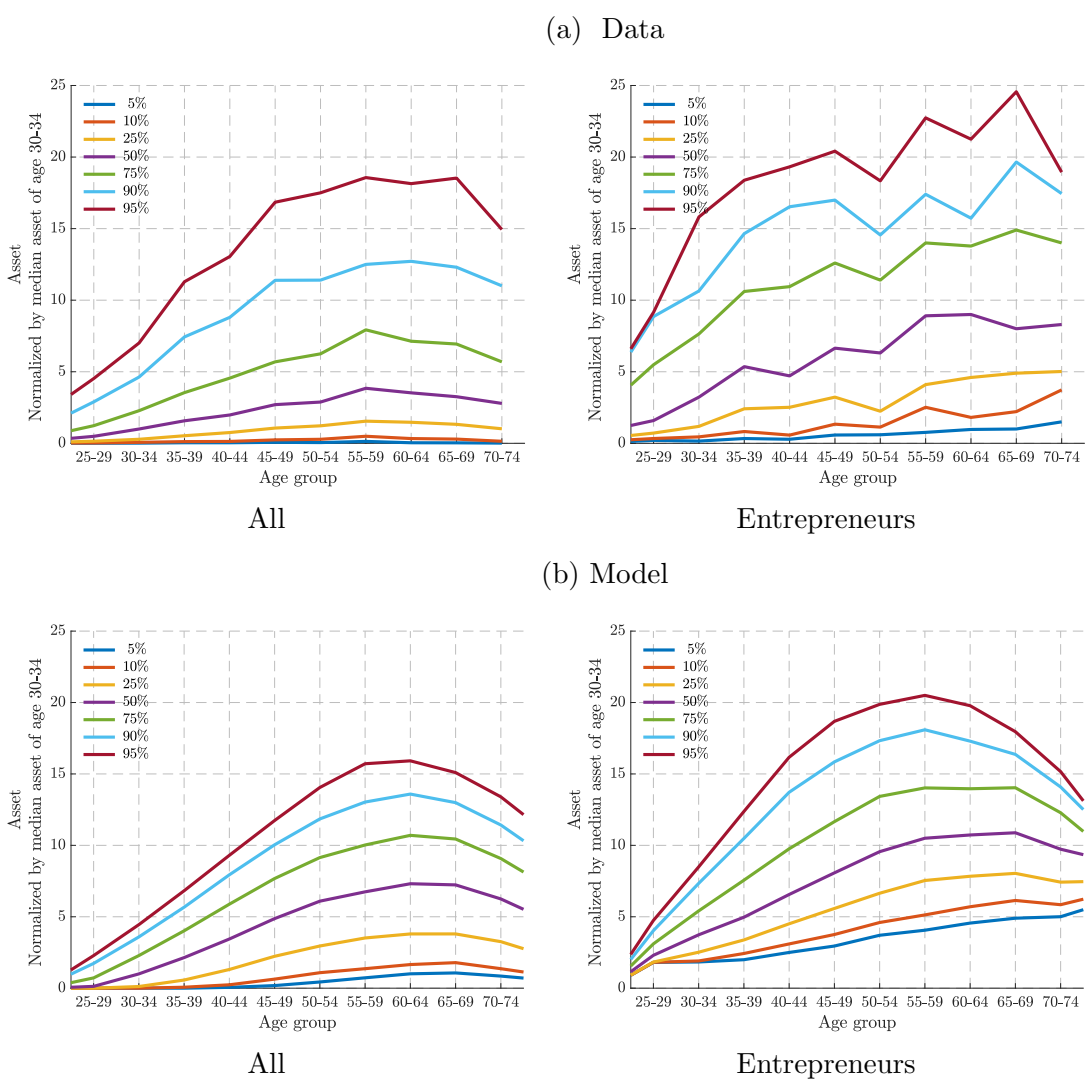

Figure A12: Model Fit: Asset Percentiles over the Life Cycle

the labor market. After they decide to be an entrepreneur, there will still be transitory shocks realized to their innate productivity, which essentially determines the productivity with which output is produced. Thus, our benchmark model can easily nest the perfect information case.

By comparing our baseline estimation results with those of re-estimation under perfect information, we aim to understand how the key elements of our benchmark model—uncertainty arising from imperfect information and learning—affect the parameter estimates on entrepreneurial dynamics. Particularly, we examine how learning contributes to generating a declining exit rate with respect to entrepreneurial spell or working age endogenously.

To proceed, we focus only on the parameters related to entrepreneurial dynamics while fixing the remaining ones to the benchmark estimates. This leaves us with six

parameters to re-estimate: $\{\mu_e, \nu_e, \sigma_e, \kappa_o, \kappa_e, \delta_e\}$. We jointly calibrate these parameters to match six empirical moments related to the entrepreneurial sector: (1) population share of entrepreneurs; (2) annual entry rate; (3) annual exit rate; (4) ratio of mean income of entrepreneurs to mean income of workers; (5) ratio of median income of entrepreneurs to median income of workers; and (6) fraction of entrepreneurs with negative incomes.

Table A15: Comparison of Estimates: Benchmark v.s. Perfect Information

Parameter	Description	Benchmark	Perfect Information
μ_e	Mean entrep. productivity	1.31	1.40
ν_e	Std. dev.: innate entrep. prod.	0.17	0.10
σ_e	Std. dev.: i.i.d. shocks	0.29	0.33
κ_o	Per period operational cost	0.02	0.06
κ_e	One-time entry cost	0.05	0.00
δ_e	Exogenous exit rate	0.02	0.09

The main takeaway is that in the perfect information case, where the learning channel driving entrepreneurial choice is absent, entrepreneurial exits are primarily explained by the exogenous separation shock δ_e. This can be observed in the last row of Table A15. Consequently, we do not see the pattern of declining exit rates with respect to working age, especially at young ages.

One additional result we learn from this exercise is that since young agents in our benchmark model are more likely to experiment with entrepreneurship to explore their innate ability, the majority of business losses occur at young ages, and the fraction of entrepreneurs with negative business income decreases by age. In contrast, under the case of perfect information, the fraction slightly increases by age because the entry cutoff decreases as agents accumulate assets over their lifetime.

B.4.2 Entrepreneurship over the Life Cycle by Innate Ability

In Figure A13, we report the share of entrepreneurs in the population with respect to age across various innate entrepreneurial ability types for both benchmark case and perfect information case. In line with our primary findings in the main text, under perfect information, individuals with high innate abilities have a significantly larger share of their lifetime spent as entrepreneurs. This increase is particularly pronounced from a young age.

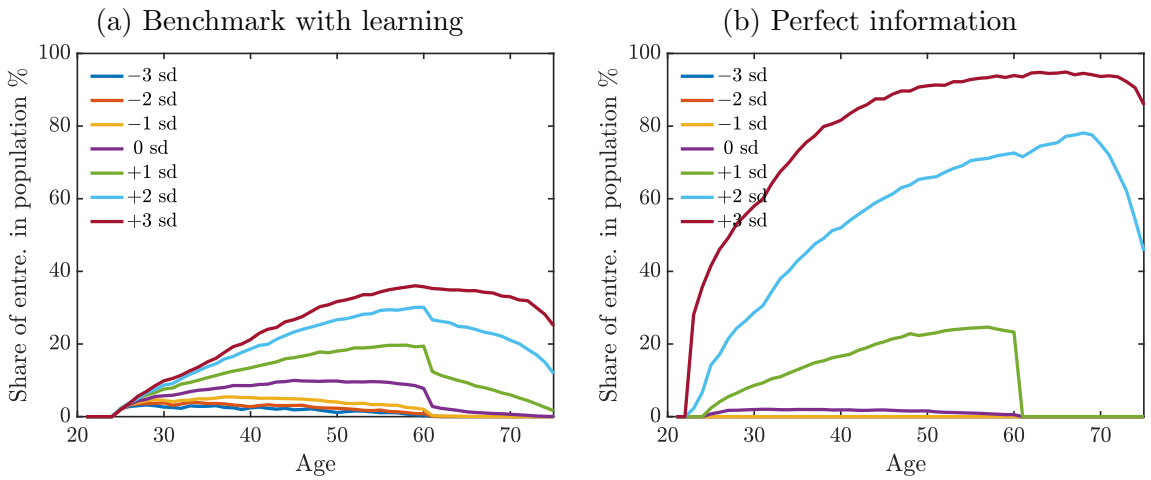

Figure A13: Entrepreneur Share by Innate Productivity Type over the Life Cycle

B.4.3 Interaction between Informational and Financial Frictions

Besides the saving behavior for workers and entrepreneurs in both benchmark and perfect information scenarios reported in the main text, we also investigate the saving behavior of various entrepreneurial ability types in these scenarios. As illustrated in Figure A14, low-ability-type agents display similar hump-shaped saving patterns throughout their life cycle in both cases. In the perfect information scenario, high-ability-type agents save substantially from the very beginning. However, in the benchmark case, due to limited information on their ability types, these agents save less during their early years compared to the perfect information scenario.

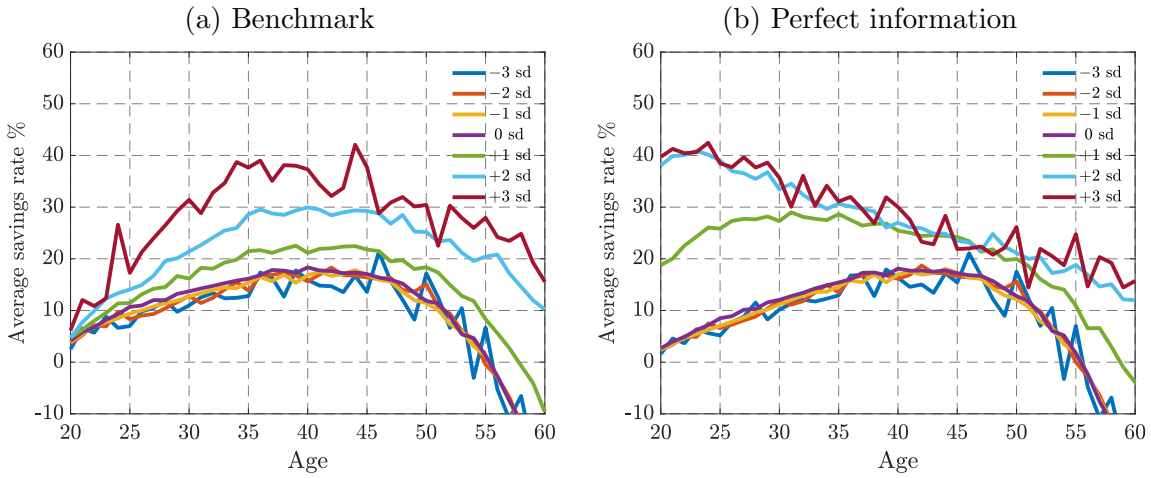

Figure A14: Saving Rates by Innate Entrepreneurial Productivity Types

Next, to help further understand Panel (B) of Figure 9 in the main text, we report the

percentage point change in both entry rate and exit rate after the collateral constraint is eased, i.e., λ is raised from 1.5 to 2, in Figure A15.

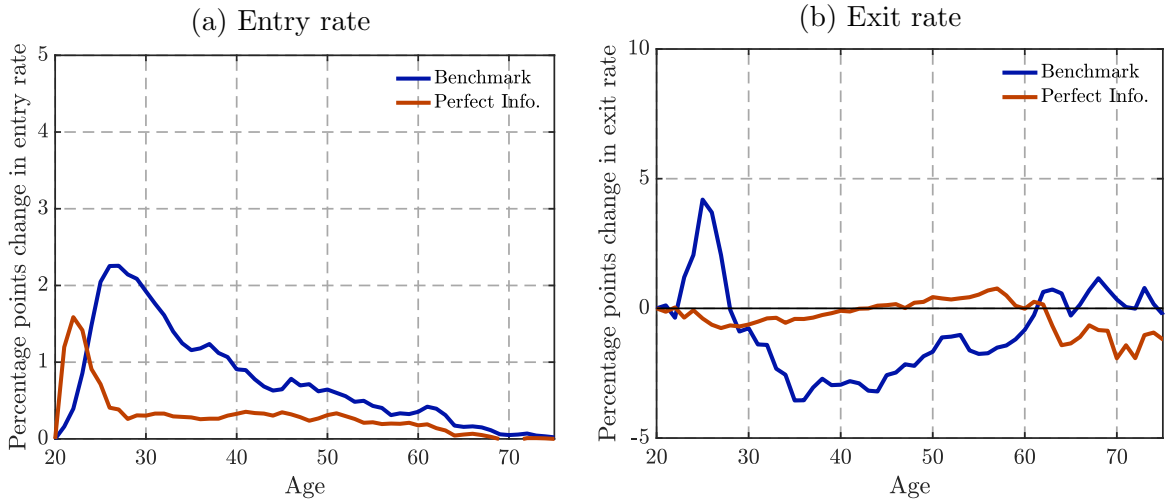

Figure A15: Percentage Point Change before and after the Collateral Constraint is Relaxed

In Table A16, we provide further analysis on the impact of relaxing the collateral constraint on aggregate moments. We report impact on moments when prices and parameters are fixed. Panel A presents data on the life cycle entry and exit of entrepreneurs, as well as the fractions of lending entrepreneurs and entrepreneurs facing constraints. Panel B focuses on aggregate capital and output. As shown in Table A16, under perfect information, where individuals perfectly know their comparative advantage since entering the labor market, the relaxation of the collateral constraint has a much smaller effect on life cycle entrepreneurship, entry and exit rates. Note that, as demonstrated in Table A15, the recalibration will lead to a lower spread of innate ability dispersion, thus the increase in population share of entrepreneurs and average output per entrepreneur would be further attenuated in a recalibrated version.

B.5 Additional Results of Policy Impacts

In Table A17, we present the effects of the revenue-neutral flat business income tax reform on aggregate moments for both the benchmark and perfect information cases. In the benchmark case, which includes informational frictions and learning, the flat tax reform significantly reduces the aggregate share of entrepreneurs and entrepreneurial output, as high innate entrepreneurial ability agents are most affected by this tax policy change. As a result, total output also decreases. However, in the perfect information case, the redistribution effects benefit agents with the highest entrepreneurial ability, despite a

Table A16: Impact on Aggregate Moments When the Collateral Constraint is Relaxed

	Benchmark with learning			Perfect information		
	$\lambda = 1.5$	$\lambda = 2.0$	Δ, p.p.	$\lambda = 1.5$	$\lambda = 2.0$	Δ, p.p.
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
Panel A. Entry/exit/collateral-constrained/lending of entrep.						
Entrep. pop. share	0.0969	0.1430	4.61	0.1352	0.1607	2.55
Entry rate	0.0204	0.0306	1.02	0.0165	0.0197	0.32
Exit rate	0.2207	0.2048	-1.59	0.1111	0.1087	-0.24
Frac. of constrained	0.6377	0.5652	-7.25	0.6835	0.5906	-9.29
Frac. of lending	0.7286	0.7249	-0.37	0.7707	0.7507	-2.00
	$\lambda = 1.5$	$\lambda = 2.0$	Δ,%	$\lambda = 1.5$	$\lambda = 2.0$	Δ,%
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
Panel B. Aggregate capital and output (normalized)						
Capital, aggregate	1	1.1444	14.44	1.2273	1.3429	9.42
Capital, entre. production	1	1.8901	89.01	1.8942	2.7854	47.04
Capital, lending	1	2.7965	179.65	1.9357	4.2220	118.11
Output, aggregate	1	1.1322	13.22	1.1624	1.2490	7.45
Output, entre.	1	1.7361	73.61	1.9719	2.6473	34.25
Output per entre	1	1.1307	13.07	1.3349	1.4748	10.48

slight decrease in the aggregate share of entrepreneurs. Consequently, the decline in entrepreneurial output is minimal, and total output increases.

Table A17: Impacts of the Revenue-neutral Flat Tax Reform on Aggregate Moments

	Benchmark with learning	Perfect information
Self-employment rate	–36.3%	–16.3%
Interest rate	4.7%	-5.0%
Wage rate	–1.1%	0.9%
Total output	–1.6%	1.8%
Private business	–26.5%	-1.4%
Coporate	16.5%	10.4%
Ave. private business output	16.1%	18.1%
Agg. employee hours	1.0%	1.3%
Agg. capital	5.5%	10.3%
AMTR-worker	1.4%	0.9%
AMTR-entre.	–46.3%	–45.6%
ATR-worker	0.8%	1.0%
ATR-entre.	-15.1%	–12.7%

References

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

Dillon, Eleanor W. and Christopher T. Stanton. 2018. "Self-employment Dynamics and the Returns to Entrepreneurship." Working paper.

Drew, Julia A. Rivera, Sarah Flood, and John Robert Warren. 2014. "Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 months." *Journal of Economic and Social Measurement* 39 (3):121–144.

Hamilton, Barton H., Nicholas W. Papageorge, and Nidhi Pande. 2019. "The Right Stuff? Personality and Entrepreneurship." *Quantitative Economics* 10 (2):643–691.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Review of Economic Dynamics* 13 (1):15–51.

Hincapié, Andrés. 2020. "Entrepreneurship over the Life Cycle: Where are the Young Entrepreneurs?" *International Economic Review* 61 (2):617–681.

Hurst, Erik and Benjamin Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* :73–143.

Lise, Jeremy and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *American Economic Review* 110 (8):2328–2376.