Experimentation and the Returns to Entrepreneurship*

Gustavo Manso
Haas School of Business,
University of California at Berkeley

March 25, 2016

Abstract

Previous studies have argued that entrepreneurs earn less and bear more risk than salaried workers with otherwise similar characteristics. In a simple model of entrepreneurship, I show that estimates of mean and variance of returns to entrepreneurship used by these previous studies are biased, as they fail to account for the option value of experimenting with new ideas. Using longitudinal data, I find patterns that are consistent with entrepreneurship as experimentation and returns to entrepreneurship that are more attractive than established by previous research. (JEL L26, D83, J24)

*I thank Daniel Ferreira, David Levine, Ross Levine, Ulrike Malmendier, Robert Marquez, Adair Morse, Matt Rhodes-Kropf, David Sraer, Toby Stuart, and Annette Vissing-Jorgensen for helpful comments. Send correspondence to Gustavo Manso, Haas School of Business, University of California at Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720; telephone: (510)643-6623. E-mail: manso@haas.berkeley.edu.

Previous research has found that entrepreneurs earn less and bear more risk than salaried workers, raising the question of why people choose to become entrepreneurs (Hamilton 2000; Moskowitz and Vissing-Jorgensen 2002). Prevailing explanations for this puzzle are based on nonstandard beliefs or preferences. For example, entrepreneurs may enjoy nonpecuniary benefits (Blanchflower and Oswald 1992), may have a preference for skewness (Kraus and Litzenberger 1976), or may just be overconfident (Cooper et al. 1988; Arabsheibani et al. 2000; Bernardo and Welch 2001).

Nonstandard beliefs or preferences may not be necessary to justify the decision to become an entrepreneur. Studies showing that entrepreneurship does not pay mostly rely on cross-sectional data to compute estimates of the mean and standard deviation of entrepreneurial earnings. I show that these estimates do not reflect the actual risk and return that individuals face when they decide to become entrepreneurs as they fail to account for the option value of experimenting with new ideas.

Using longitudinal data, I find patterns that are consistent with entrepreneurship as experimentation: entrepreneurship spells are short; the probability of abandoning entrepreneurship is higher after bad performance; and failed entrepreneurs are not punished when they return to the salaried workforce.

Lifetime earnings computed from longitudinal data incorporate the value of the options embedded in entrepreneurship. Once the value of these options are taken into account returns to entrepreneurship are more attractive than suggested by previous research. Successful entrepreneurs earn significantly more than salaried workers with similar characteristics, while failed entrepreneurs are able to quickly move back to the salaried workforce limiting their losses. The option to abandon entrepreneurship increases the return and reduces the risk faced by entrepreneurs.

This view of entrepreneurship as experimentation and real options flips the interpretations of some of the previous findings. High variance in cross-sectional self-employed earnings, as found in previous research, is actually valuable for entrepreneurs since this variance increases the value of their real options. Failed entrepreneurs will quickly abandon entrepreneurship and variance in cross-sectional earnings will not be reflected in lifetime earnings.

To study the distribution of entrepreneurial returns, I develop a simple model of entrepreneurship as experimentation with new ideas. In the model, individuals with new ideas may pursue

them as self-employed workers, and this is the only way to find out whether an idea is good. Alternatively, individuals may remain salaried workers.

The model reveals how cross-sectional data analysis can introduce different sources of bias in estimating the distribution of entrepreneurial returns. "Survivorship bias" arises because the cross-sectional distribution overweights successful entrepreneurs who survive longer. "Experimentation bias" arises because the cross-sectional distribution neglects the fact that entrepreneurs who fail will not carry on with their bad ideas, but instead will switch to being salaried workers or trying new ventures.

These biases affect estimates of the mean and the variance of entrepreneurial returns. Survivorship bias leads to an overstatement of the true lifetime mean of self-employed earnings, while the experimentation bias leads to an understatement of the true lifetime mean of self-employed earnings. Depending on which effect dominates, the cross-sectional mean of self-employed earnings may overstate or understate the lifetime mean of self-employed earnings. On the other hand, since the experimentation bias amplifies entrepreneurial failures and the survivorship bias overweights successful entrepreneurs, the cross-sectional variance of self-employed earnings typically overstates the lifetime variance of self-employed earnings.

An extension of the model studies what happens if previous entrepreneurial experience generates an earnings premium for salaried workers. In such settings, cross-sectional data analysis introduces a new source of bias. "Attribution bias" arises because the cross-sectional distribution of salaried earnings fails to account for the fact that the wage premium earned by salaried workers may be a consequence of previous entrepreneurial experience. Attribution bias will make the cross-sectional mean earnings of salaried workers overstate the lifetime mean earnings of salaried workers, while it will make the cross-sectional mean earnings of self-employed workers understate the cross-sectional mean earnings of self-employed workers.

To test the predictions of the model, I use the National Longitudinal Survey of Youth-1979 (NLSY79). From the NLSY79, I obtain information on demographics, educational attainment, labor market outcomes, and prelabor market traits. The main advantage of the NLSY79 is that it follows individuals over time, allowing one to compute the lifetime returns to self-employed and

salaried workers.[1]

According to the model, self-employed workers experiment with new ideas when they leave the salaried workforce. The value of experimentation arises from the option to abandon bad ideas. For this option to be valuable, self-employment spells must be short, particularly for workers who perform poorly as self-employed. I find that approximately 52% of entrepreneurship spells in NLSY79 lasts less than two years. Moreover, a probit regression estimating how residual earnings affect the probability of abandoning entrepreneurship shows that lower residual earnings while self-employed are associated with a higher probability of abandoning self-employment.

It is also important for experimentation to be valuable that individuals are not penalized for previous entrepreneurial failures. I find that salaried workers earn a premium if they have previously completed a self-employment spell. This shows that the option to abandon self-employment is there and is attractive for the self-employed.

To study lifetime returns to entrepreneurship, I divide the sample into two groups: those who were ever self-employed and those who were never self-employed. Mean lifetime earnings of the ever self-employed are higher than mean lifetime earnings of the never self-employed. More importantly, in contrast to previous studies that rely on cross-sectional data, the risk as measured by the standard deviation of earnings is not substantially greater than the risk of staying a salaried worker.

The comparison between lifetime earnings of ever self-employed and never self-employed has important shortcomings. If an individual enters self-employment late in life, she is classified as ever self-employed, even though most of her earnings come from being a salaried worker without any entrepreneurial experience.

To address this issue, I use propensity score matching to compare the earnings of an individual who chooses to become self-employed with someone who looks just like this individual in terms of observed characteristics but decides to remain a salaried worker. I find that, on average, after becoming self-employed, individuals earn approximately 5% less during the first couple of years, but earn approximately 10% more than their salaried counterparts in the subsequent years.

Conditioning the analysis on the number of years as self-employed, I find that individuals who

[1]Fairlie (2005) and Levine and Rubinstein (2015) are examples of studies that use these data.

attempt to be entrepreneurs but abandon self-employment in less than two years are not punished, achieving approximately the same earnings as those who have not attempted self-employment. Individuals who stay self-employed longer than two years earn substantially higher earnings than salaried workers with similar characteristics.

The model of entrepreneurship as experimentation used in this paper follows a long tradition in the study of entrepreneurship and innovation. Schumpeter (1934) argues that entrepreneurship is essentially experimentation with "new combinations" of existing resources. Arrow (1969) associates innovation with the production of knowledge through experimentation. March (1991) uses the terms exploration and exploitation to describe the fundamental tension that arises in learning through experimentation.

Manso (2011) studies incentives for exploration and exploitation. It shows that tolerance for early failure and reward for long-term success are optimal to motivate exploration and, consequently, entry into entrepreneurship. As the paper argues, such incentive scheme can be implemented with debtor-friendly bankruptcy laws. Other papers have focused on institutional aspects of the labor market that also offer protection against failure and thus motivate entrepreneurship, such as unemployment insurance (Hombert et al. 2015) and job-protected leave (Gottlieb et al. 2016). Along these lines, the equilibrium in the labor market uncovered here, which presents no stigma for failed entrepreneurs, should serve to encourage entrepreneurship.

In this paper, I focus on showing that previous estimates of the returns to entrepreneurship are biased because they fail to take into account the option value of experimenting with new ideas. Other papers try to resolve the puzzle on returns to entrepreneurship by providing behavioral explanations for why entrepreneurs might accept to work for less, such as risk preferences, overconfidence, and nonpecuniary benefits. Astebro et al. (2014) provide a survey of this literature and conclude that behavioral research has not yet provided definitive explanations for the puzzling aspects of entrepreneurship.

More closely related, Levine and Rubinstein (2015) argue that the puzzle may be due to mismeasurement because self-employment is not a good measure of entrepreneurship. They show that self-employed workers who incorporate their business earn substantially more than salaried workers and argue that only self-employed workers who incorporate should be called entrepreneurs.

The paper is also related to dynamic models of discrete occupational choices. These models have been successful in explaining issues such as patterns of wealth distribution, the role of financial intermediaries, and the effects of changes in the tax or bankruptcy regulation. Cagetti and De Nardi (2006), Hintermaier and Steinberger (2005), Vereshchagina and Hopenhayn (2009), Campanale (2010), and Poschke (2013) develop dynamic models of occupational choice in which workers can choose to become entrepreneurs and learn about their entrepreneurial skills. However, these models focus on different questions and do not directly compare cross-sectional and lifetime entrepreneurial earnings.

1. A Model of Entrepreneurship as Experimentation

This section introduces a simple overlapping-generations model to study the returns to entrepreneurship. In each period $t \in \{0, 1, \ldots\}$, a unit mass of agents is born. All agents live for two periods and are risk neutral with zero discounting.

When born, a fraction γ of agents have access to new ideas, which they may pursue as self-employed workers. Alternatively, agents may work as salaried workers. As salaried workers, agents receive a wage W each period. If an agent has an idea and chooses to pursue it as self-employed, he may find the idea is high quality with probability p, in which case it pays out R each period, or low quality with probability $1-p$, in which case it pays out 0. The only way to find the quality of a new idea is by trying it out as a self-employed worker.

To capture the exploratory nature of self-employment, I assume that $R > W$ and $pR < W$. If successful, the self-employed earn more than salaried workers. However, the unconditional mean of self-employed earnings is lower than salaried earnings.

Under these assumptions, there are two strategies that need to be considered. Agents may choose to always remain salaried workers, earning $V_{\text{sal}} = W$ per period. Alternatively, agents may become self-employed if they have a new idea. They will remain self-employed if their idea is of high quality, since it yields $R > W$ in each period. If their idea is low quality, it yields 0, and they will abandon it and return to the salaried workforce. The expected per period earnings

V_{semp} of such strategy are:

$$V_{\text{semp}} = pR + (1-p)\frac{W}{2}. \tag{1}$$

The intuition for Equation (1) is as follows. Self-employed workers have a high-quality idea with probability p, in which case they earn R each period. With probability $1-p$ they have a low-quality idea, in which case they earn zero for one period and become a salaried worker thereafter, earning W in the second period.

Agents earn more as self-employed than as a salaried worker if and only if $V_{\text{semp}} \geq V_{\text{sal}}$, which is equivalent to

$$pR \geq \frac{(1+p)}{2}W. \tag{2}$$

Otherwise, agents earn more being a salaried workers.

The above comparison only takes into account monetary payoffs of workers. To ensure that agents with ideas try them even if monetary payoffs are not enough to justify becoming self-employed, we assume that self-employed workers enjoy private benefits β that are high enough to make entrepreneurship pay off. In the model of this section, this condition holds if $\beta \geq W - 2pR/(1+p)$. This assumption will play no other role in the analysis as we will focus on monetary payoffs, which are observable.

To understand cross-sectional data generated by the model, we can calculate the distribution of agents in the population at any time t. Let θ_{sal}, $\theta_{\text{semp,f}}$, and $\theta_{\text{semp,s}}$ be the fractions of salaried workers, successful self-employed workers, and failed self-employed workers in the population at any point in time. These fractions are given by:

$$\begin{aligned} \theta_{\text{sal}} &= (1-\gamma) + \frac{\gamma(1-p)}{2}, \\ \theta_{\text{semp,f}} &= \frac{\gamma(1-p)}{2}, \\ \theta_{\text{semp,s}} &= \gamma p. \end{aligned} \tag{3}$$

The first fraction in (3), θ_{sal}, consists of $(1-\gamma)$ young and old individuals who are not born with an idea and $\gamma(1-p)/2$ old individuals who were born with a bad idea. The second fraction, $\theta_{\text{semp,f}}$, consists of $\gamma(1-p)/2$ young individuals who were born with a bad idea. The third fraction, $\theta_{\text{semp,s}}$, consists of γp young and old individuals who were born with a good idea.

Using (3), we can compute cross-sectional earnings distributions and compare those with lifetime earnings distributions. Cross-sectional data introduces two sources of bias in estimating lifetime earnings distributions for self-employed workers. Survivorship bias arises because the cross-sectional distribution overweights successful self-employed workers who survive longer as self-employed. Experimentation bias arises because the cross-sectional distribution neglects the fact that self-employed workers who fail will not carry on with their bad ideas, but instead will switch to being a salaried worker or try new ideas.

Figure 1 shows the cross-sectional and lifetime distributions of earnings for salaried and self-employed workers for the model with the following parameters: $W = 30, R = 60, p = 0.4, \gamma = 0.05$. These distributions illustrate the survivorship and experimentation biases. Because of survivorship bias, in the cross-sectional earnings distribution, the probability of being successful as a self-employed worker is higher than in the lifetime distribution. At the same time, due to experimentation bias, in the cross-sectional earnings distribution, the probability of failing and earning 0 is higher than in the lifetime earnings distribution. The lifetime earnings distribution correctly reflects the fact that these failed self-employed workers will earn zero just for one period and will switch to the salaried workforce earning a lifetime mean payoff that is between zero and W.

[**Figure** 1 about here.]

The next proposition compares the cross-sectional mean of self-employed earnings with the lifetime mean of self-employed earnings.

Proposition 1 The cross-sectional mean of self-employed earnings overstates the lifetime mean of self-employed earnings if and only if the lifetime mean of self-employed earnings is higher than salaried workers wage W.

Proof The cross-sectional mean of self-employed earnings is:

$$V_{\text{cs}} \equiv \frac{\theta_{\text{semp,s}}}{\theta_{\text{semp,s}} + \theta_{\text{semp,f}}} R = \frac{2p}{1+p} R, \tag{4}$$

which is greater than V_{semp} if

$$\frac{2p}{1+p} R \geq pR + \frac{(1-p)W}{2} \;\Leftrightarrow\; pR \geq \frac{(1+p)}{2} W. \tag{5}$$

■

Survivorship bias leads to an overstatement of the true lifetime mean of self-employed earnings. Experimentation bias amplifies entrepreneurial failure, leading to an understatement of the true lifetime mean of self-employed earnings. If the lifetime mean of self-employed earnings is higher than the lifetime mean of salaried earnings, then survivorship bias prevails and the cross-sectional mean of self-employed earnings overstates the lifetime mean of self-employed earnings. Otherwise, the opposite holds.

The next proposition compares the cross-sectional standard deviation of salaried and self-employed earnings with the lifetime standard deviation of salaried and self-employed earnings.

Proposition 2 The following statements about standard deviations of earnings hold:

1. The cross-sectional standard deviation of salaried earnings is equal to the lifetime standard deviation of salaried earnings.

2. There exists $\overline{\lambda} \in (1, \infty]$ such that the cross-sectional standard deviation of self-employed earnings overstates the lifetime standard deviation of self-employed earnings if and only if $\lambda \equiv V_{\text{semp}}/V_{\text{sal}} < \overline{\lambda}$.

Proof On point (1), both the cross-section standard deviation of salaried earnings and the lifetime standard deviation of salaried earnings are zero.

On point (2), the cross-sectional variance of self-employed earnings is:

$$\left(\frac{2p}{1+p}\right)\left(1-\frac{2p}{1+p}\right)R^2 = p(1-p)\frac{2}{(1+p)^2}R^2. \tag{6}$$

The lifetime variance of self-employed earnings is:

$$p(1-p)\left(R-\frac{W}{2}\right)^2 = p(1-p)\left(\frac{2\lambda-1}{2\lambda-(1-p)}\right)^2 R^2, \tag{7}$$

which is increasing in λ and lower than (6) if $\lambda = 1$ (or $V_{\text{semp}} = V_{\text{sal}}$). ■

Salaried work is always rewarded with W and is thus risk-free. Both the cross-sectional and lifetime standard deviation of earnings reflect that and are equal to zero.

The second statement is more subtle. The experimentation bias amplifies entrepreneurial failures, while the survivorship bias overweights successful entrepreneurs, in principle contributing

to an overstatement of the variance of self-employed earnings. It is only for extreme cases in which lifetime self-employement payoff is substantially higher than salaried payoff that the cross-sectional variance understates the lifetime variance of self-employed earnings. These situations are unlikely to arise since, in a general equilibrium model, these payoffs have to be close to each other as any substantial difference between the payoffs would attract more people to entrepreneurship.

2. Prior Entrepreneurial Experience

In this section we consider the same model as in the previous section, except that after experiencing an entrepreneurial failure, salaried workers earn κW. When $\kappa > 1$, failed self-employed workers earn a premium in the job market. When $\kappa < 1$, failed self-employed workers earn a discount in the job market. When $\kappa = 1$, we return to the setup studied in the previous section.

Self-employed lifetime mean earnings are:

$$V_{\text{semp}} = pR + (1-p)\frac{\kappa W}{2}, \tag{8}$$

while the lifetime mean earnings of salary workers are $V_{\text{sal}} = W$.

The cross-sectional distribution needs to account for fractions of successful self-employed, failed self-employed, and salaried workers with and without a previous entrepreneurial failure. These are represented by $\theta_{\text{semp,s}}$, $\theta_{\text{semp,f}}$, $\theta_{\text{sal}}(1)$, and $\theta_{\text{sal}}(0)$, respectively. The cross-sectional distribution is given by:

$$\begin{aligned} \theta_{\text{semp,s}} &= \gamma p, \\ \theta_{\text{semp,f}} &= \frac{\gamma(1-p)}{2}, \\ \theta_{\text{sal}}(1) &= \frac{\gamma(1-p)}{2}, \\ \theta_{\text{sal}}(0) &= (1-\gamma). \end{aligned} \tag{9}$$

The first fraction in (9), $\theta_{\text{semp,s}}$, consists of γp young and old individuals who were born with a good idea. The second fraction, $\theta_{\text{semp,f}}$, consists of young individuals who were born with a bad idea. The third fraction, $\theta_{\text{sal}}(1)$, consists of $\gamma(1-p)/2$ old individuals who pursued a bad idea, failed as self-employed and are now salaried workers. The last fraction, $\theta_{\text{sal}}(0)$, consists of $(1-\gamma)$ young and old individuals who have not pursued an idea.

Proposition 3 The cross-sectional mean of salaried workers earnings overstates the lifetime mean of salaried workers earnings if and only if previous entrepreneurial failures improve salaried worker earnings.

Proof The cross-sectional mean of salaried workers earnings is given by:

$$\frac{\theta_{\text{sal}}(0)W + \theta_{\text{sal}}(1)(\kappa W)}{\theta_{\text{sal}}(0) + \theta_{\text{sal}}(1)} = \frac{2(1-\gamma) + \gamma(1-p)\kappa}{2(1-\gamma) + \gamma(1-p)}W, \tag{10}$$

which is greater than W if and only if $\kappa > 1$. ■

If previous entrepreneurial failures improve salaried worker earnings ($\kappa > 1$), it means that salaried workers will only have access to this wage premium if they were previously self-employed. However, this wage premium is reflected in the cross-sectional distribution of salaried earnings, as no distinction is made whether or not the worker has formerly been self-employed. Because of this attribution bias, the cross-sectional mean earnings of salaried workers overstates the lifetime mean of salaried workers earnings.

Proposition 4 There exists $\overline{\kappa}$ such that for $\kappa > \overline{\kappa}$ the cross-sectional mean of self-employed earnings understates the lifetime mean of self-employed earnings.

Proof The cross-sectional mean of self-employed earnings is:

$$V_{\text{cs}} \equiv \frac{\theta_{\text{semp,s}}}{\theta_{\text{semp,s}} + \theta_{\text{semp,f}}}R = \frac{2p}{1+p}R, \tag{11}$$

which is independent of κ. However, it is clear from Equation (8) that V_{semp} is increasing in κ and goes to infinity as κ goes to infinity. ■

The cross-sectional distribution of self-employed earnings does not take into account the salaried wage premium κ that only accrues to the worker if he has previous experience as self-employed. For high enough κ, this attribution bias becomes dominant and the cross-sectional mean of self-employed earnings understates the lifetime mean of self-employed earnings.

Another form of attribution bias can arise if successful self-employed workers become a salaried worker (manager) in their own firms. In this case, cross-sectional studies would count them as salaried workers, but their high earnings are there only because they were entrepreneurs in the first place. Our empirical approach, using longitudinal data will deal with this issue.

We now turn to the standard deviation of salaried and self-employed earnings in the presence of attribution bias.

Proposition 5 The following statements about standard deviations of earnings hold:

1. The cross-sectional standard deviation of salaried earnings overstates the lifetime standard deviation of salaried earnings if and only if $\kappa \neq 1$.

2. For κ close to 1, there exists $\overline{\lambda} \in (1, \infty]$ such that the cross-sectional standard deviation of self-employed earnings overstates the lifetime standard deviation of self-employed earnings if and only if $\lambda \equiv V_{\text{semp}}/V_{\text{sal}} < \overline{\lambda}$.

Proof On point (1), the lifetime standard deviation of salaried earnings is zero. The cross-sectional standard deviation of salaried earnings is greater than zero if and only if $\kappa \neq 1$ as the cross-sectional distribution of salaried earnings include previously failed entrepreneurs making κW.

On point (2), the cross-sectional variance of self-employed earnings is:

$$\left(\frac{2p}{1+p}\right)\left(1-\frac{2p}{1+p}\right)R^2 = p(1-p)\frac{2}{(1+p)^2}R^2. \tag{12}$$

The lifetime variance of self-employed earnings is:

$$p(1-p)(R-\frac{W}{2})^2 = p(1-p)\left(\frac{2\lambda-\kappa}{2\lambda-(1-p)\kappa}\right)^2 R^2, \tag{13}$$

which is increasing in λ and lower than (12) if $\lambda = 1$ (or $V_{\text{semp}} = V_{\text{sal}}$) and κ is close to 1. ■

With attribution bias, the cross-sectional standard deviation of salaried earnings is greater than zero and thus overestimates the lifetime standard deviation of salaried earnings. As before, the cross-sectional standard deviation of self-employed earnings typically overestimates the lifetime standard deviation of self-employed earnings.

3. The Returns to Self-Employment: Evidence from the NLSY79

This section examines the returns to salaried and self-employed workers using the National Longitudinal Survey of Youth-1979 (NLSY79). The main advantage of the NLSY79 is that it follows individuals over time, allowing one to compute the lifetime returns to self-employment.

3.1 Data

The NLSY79 is a survey of 12,686 individuals who were 15-22 years old when first surveyed in 1979. The same individuals were then surveyed annually through 1994 and every two years thereafter. The last survey year in my sample is 2012. To keep the frequency of observations constant throughout the sample period, I drop odd years from the sample. I also drop from the sample observations in which an individual has missed the survey.

The NLSY79 cohort is comprised of three subsamples. For this study, I drop the military and representative minorities subsamples to restrict the analysis to the nationally representative subsample of 6,111 individuals. I also dropped from the sample all observations corresponding to individuals who were never a worker during the period 1979-2012. The final sample contains observations on 5,415 individuals.

From the NLSY79, I obtain information on demographics, educational attainment, labor market outcomes, and prelabor market traits. The demographic variables are age, gender, and race. For educational attainment, I construct dummy variables for six education categories based on years of schooling. Labor market outcomes include earnings, hours worked, weeks worked, and industry.

Prelabor market traits include three different measures. To measure cognitive ability, I use the Armed Forces Qualifications Test (AFQT) score, which measures the aptitude and trainability of each individual. Collected during the 1980 NLSY79 survey, the AFQT score is based on information concerning arithmetic reasoning, world knowledge, paragraph comprehension, and numerical operations. It is frequently employed as a general indicator of cognitive skills and learning aptitude. The AFQT score is measured as a percentile of the NLSY79 survey, with a median value of 50.

To measure self-esteem, I use the Rosenberg Self-Esteem score, which is based on a ten-part questionnaire given to all NLSY79 participants in 1980. It measures the degree of approval or disapproval of one's self. The values range from six to 30, where higher values are associated with greater self-approval.

I also use information on the degree to which individuals believe they have internal control of their lives through self-determination relative to the degree that external factors, such as chance,

fate, and luck, shape their lives. This is measured by the Rotter Locus of Control, which was collected as part of a psychometric test in the 1979 NLSY79 survey. The Rotter Locus of Control ranges from four to sixteen, where higher values signify less internal control and more external control.

All earnings variables are in thousands of dollars. They are adjusted for inflation using the Consumer Price Index (CPI) and expressed in 2012 dollars.

3.2 Pooled data

In this subsection, I take individual-year observations as the basic unit of analysis, to provide summary statistics and also reproduce results obtained in previous studies using cross-sectional data.

Table 1 shows summary statistics from NLSY79 data. Self-employed workers are similar to salaried workers in most characteristics, but typically there is a higher proportion of white and males among the self-employed. Mean annual earnings ($\$52,250$) of self-employed workers are higher than those ($\$47,040$) of salaried workers. However, median annual earnings ($\$31,130$) of self-employed workers are lower than those ($\$37.36$) of salaried workers. Moreover, the standard deviation ($\$73,120$) of self-employed workers' earnings is substantially higher than that ($\$45,750$) of salaried workers' earnings. These are in line with previous studies that conclude that the median self-employed individual earns less and bears significantly more risk than salaried workers.[2] Even though mean self-employed earnings are higher than salaried workers earnings, it is hard to justify selection into entrepreneurship given the difference in risk between the two career choices as captured by the standard deviation of earnings. Entrepreneurship seems too risky as the standard deviation of self-employed workers' earnings is substantially higher than the standard deviation of salaried workers' earnings.

[**Table** 1 about here.]

[2]See, for example, Table 3 in Hamilton (2000) or Figure 2 of Moskowitz and Vissing-Jorgensen (2002).

3.3 Lifetime earnings

This subsection exploits the longitudinal dimension of the data to compute unbiased estimates of the returns to self-employment.

According to the model in Section 1., self-employed workers are engaging in experimentation when they choose self-employment. As such, the average spell of self-employment should be short. If it takes too long to learn about the quality of an idea, there is little value in experimentation. Moreover, self-employed workers should be more likely to leave self-employment after lower earnings while self-employed.

Figure 2 shows a histogram with the duration of self-employment spells that start between 1979 and 2002. Approximately 52% of self-employment spells last less than 2 years. This is consistent with the view that the self-employed experiment with new ideas and learn quickly about the quality of their ideas. Therefore, the losses due to an entrepreneurial failure do not impose a large penalty on lifetime earnings.

[**Figure** 2 about here.]

Table 2 presents coefficient estimates for a probit model estimating how residual earnings affect the probability of abandoning self-employment. Residual earnings are the residuals of the ordinary least-squares (OLS) regression that controls for demographics, educational attainment, work experience, industry, and prelabor market traits. According to Table 2, lower residual earnings while self-employed are associated with a higher probability of abandoning self-employment. When evaluated at the means, a one-standard-deviation increase in residual earnings raises the probability of abandoning entrepreneurship from 30.4% to 35.8%, while a one-standard-deviation decrease in residual earnings reduces the probability of abandoning entrepreneurship from 30.4% to 25.5%.

[**Table** 2 about here.]

The basic model of Section 1. also predicts that self-employed workers are not penalized if they decide to return to the salaried workforce. Table 3 presents results of an OLS model estimating how a previously completed self-employment spell affects annual earnings for individuals

between 25 and 55 years old working full-time, year round. These regressions control for work experience, as well as year and individual fixed effects. Salaried workers with previous self-employment experience earn a premium of $\$5,161$ per year when compared to similar workers without self-employment experience.[3] Self-employed workers who have previously completed a self-employment spell earn a negative (but not statistically significant) premium.

These findings are consistent with the extension of the basic model considered in Section 2. with a premium for salaried workers with previous self-employment experience ($\kappa > 1$). They also suggest we should be specially careful when using cross-sectional data to measure returns to entrepreneurship since in the cross-section this self-employment premium would be attributed to salary workers, but is only attainable by entrepreneurs.

[**Table** 3 about here.]

Using NLSY79, we can compute lifetime mean earnings for different employment types. Table 4 provides summary statistics and is the analogous of Table 1 for longitudinal data. To study whether it pays off to be an entrepreneur, I classify individuals into two groups: those who have never been self-employed and those who have ever been self-employed. Summary statistics in Table 4 show a much more balanced picture of the choice between self-employment or salaried work than those in Table 1. The mean lifetime earnings ($\$42,770$) of those who were ever self-employed are slightly higher than the mean lifetime earnings ($\$40,460$) of those who were never self-employed. The median lifetime earnings ($\$34,670$) of those who were ever self-employed are slightly lower than the median lifetime earnings ($\$36,110$) of those who were never self-employed. The risk dimension is the main departure from the statistics in Table 1. Different from the results of the pooled data analysis, the standard deviation ($\$41,790$) of lifetime earnings of workers who were ever self-employed is only a bit higher than the standard deviation ($\$34,480$) of lifetime earnings of workers who have never been self-employed.

[**Table** 4 about here.]

The results in Table 4 call into question previous findings that claim that entrepreneurs earn less and bear significantly more risk than salaried workers (Hamilton 2000; Moskowitz and Vissing-

[3] Hamilton (2000) presents similar findings.

Jorgensen 2002). Once lifetime earnings are taken into account, returns of entrepreneurs appear to be higher than those of salaried workers, while entrepreneurs bear only a little more risk than salaried workers.

3.4 Propensity score matching

The comparison in the previous section was between never self-employed and ever self-employed individuals. This analysis has important shortcomings. For example, people who become entrepreneurs late in life have earnings that are counted for the ever self-employed, while most of their earnings are coming before they became self-employed.

A more precise exercise is to compare earnings of an individual who chooses to become self-employed for the first time with someone who has never been self-employed and looks just like this individual in terms of observed characteristics, but decides to remain a salaried worker.[4] An individual is considered a match to the individual that chooses self-employment if right before self-employment: (1) their earnings are in the same percentile, (2) their earnings growth are in the same decile, (3) their propensity scores based on work experience, demographics, educational attainment, prelabor market traits, industry, and year are in the same decile. If more than one individual is considered a match, I use the mean of their earnings for the analysis.

Table 5 shows the means for the treatment and control groups after matching. Two-sided t-tests indicate no significant differences at the 95% confidence level for each variable, highlighting the quality of the match.

[**Table** 5 about here.]

Figure 3 compares outcomes between the treatment and control groups. As shown in the left graph, after becoming self-employed, individuals go through a couple of years with lower earnings but then in the subsequent years they earn on average approximately 10% more per year than similar individuals who decided to remain salaried workers. The right graph shows that these differences are statistically significant in many of the years.

[**Figure** 3 about here.]

[4]Results are robust to restricting matching to individuals who never become self-employed in their lifetime.

Figure 4 compares treatment and control groups conditional on the duration of entrepreneurship span. The upper-left graph shows the results for individuals whose self-employment lasts less than two years, and the upper-right graph shows the results for individuals whose self-employment last more than two years. Individuals who attempt to be entrepreneurs but abandon entrepreneurship in less than two years, are not punished, thereby approximately achieving the same earnings as similar individuals who have not attempted to be entrepreneurs. At the same time, entrepreneurs who stay longer than two years, make substantially more than similar salaried workers.

The lower graphs in Figure 4 show the difference between the mean earnings of individuals in the treatment and control groups around the decision to become entrepreneurs (time 0). In the lower-left graph, the solid line shows the mean difference earnings between individuals in the treatment group who stay entrepreneurs for less than two years and their pairs. The dashed line represents 95% gross confidence intervals. The lower-right figure is analogous for individuals in the treatment group who stay as entrepreneurs for more than two years.

Figure 4 illustrates well the dynamic aspects of the gamble entrepreneurs face. If they fail as entrepreneurs, they can always abandon entrepreneurship without significant costs. If they succeed, they earn substantially more. As the figure shows, entrepreneurs who abandon self-employment in less than two years have low earnings at time 0, relative to their matched pairs who stayed as salaried workers. At subsequent times, after abandoning entrepreneurship, the performance of these individuals is not significantly different from their matched pairs who were never entrepreneurs. Entrepreneurs who decide to stay longer than two years, do not suffer as much at time 0 and perform significantly better than their matched pairs.

[**Figure** 4 about here.]

4. Additional Discussion

The model proposed in Sections 1. and 2. is a stylized model. There are several ways to enrich the model, some of which could make the biases described here less severe. However, there are also natural extensions that would make the biases more severe. For example, allowing agents to live more than two periods would exacerbate the experimentation bias as agents that fail as an

entrepreneur suffer only one period and have several periods ahead as a salaried worker to recover from this failure. The empirical implementation tests the main insights of the model but does not depend on its specific formulation.

Previous studies have reached the conclusion that entrepreneurs earn less and bear substantially more risk than salaried workers relying at least in part on cross-sectional data analysis.[5] However, they have attempted to mitigate some of the concerns raised in this paper. For example, Hamilton (2000) uses the 1984 Survey of Income and Program Paticipation (SIPP), which contains tenure in a particular job or business. He is thus able to estimate earnings profile as a function of tenure in a job or business. However, this still fails to correct for the attribution bias described in the current paper since it does not capture earnings after transitioning from self-employment to salaried work.

It might be possible to correct for some of the biases pointed out in this paper using cross-sectional data. However, to correct for all potential biases, cross-sectional surveys would need to ask more questions than typically available in such surveys. For example, one would need to ask whether and for how long an individual was an entrepreneur in the past to correct for the attribution bias. To get around this issue, this paper proposes correcting for these biases using longitudinal data. By following individuals over time, we can see whether they are experimenting and when they exercise their options, being thus able to value those.

One potential issue when comparing earnings of self-employed and salaried workers is the treatment of returns to capital. In the NLSY, respondents are likely to interpret the question on income as including both returns to labor and returns to capital. This is unlikely to pose a substantial problem, however, as most entrepreneurs do not invest large amounts of capital. In the 1992 Characteristics of Business Owners survey, 57% of small business require less than $\$5,000$ of startup capital (U.S. Bureau of the Census, 1997).

Moreover, for the years of 1985 to 1990 and 1992 to 1998, the NLSY contains variables that allows one to calculate an owner's equity value in the business. I follow the procedure described in Fairlie (2005) to calculate adjusted self-employed earnings by subtracting the opportunity cost

[5]See, for example, Figure 1 and Table 3 of Hamilton (2000) and Figure 2 of Moskowitz and Vissing-Jorgensen (2002).

of equity. As in Fairlie (2005), this adjustment does not affect results significantly, and therefore, I use total earnings in my analysis.

We acknowledge that our matching approach in Section 3.4 has an important limitation since the choice to become an entrepreneur is not a exogenous event. For instance, individuals with better future earnings potential may be more likely to become an entrepreneur. Matching on a long list of individual characteristics, as we do in this paper, helps make this concern less stringent. The fact that entrepreneurs do poorly initially, as predicted by the model of entrepreneurship as experimentation, is also comforting. Yet, in the absence of a proper source of exogenous variation in the probability to become an entrepreneur, our results may be subject to an endogeneity bias and should therefore be interpreted as descriptive more than causal.

The NLSY79 data used in the current paper are a household survey with no tax implications for respondents. In spite of that, Hurst, Li, and Pugsley (2014) argue that the self-employed underreport their income by 25% in household surveys. Correcting for this underreporting would make entrepreneurship even more attractive.

5. Conclusion

Previous studies showing that entrepreneurship does not pay fail to account for the option value of experimenting with new ideas. Using longitudinal data, I show that entrepreneurship is more attractive than suggested by these studies. Most entrepreneurs fail quickly and are able to limit their losses by returning to the salaried workforce. Few entrepreneurs succeed, but those who do earn significantly more than salaried workers with similar characteristics. Overall, I find that after the first couple of years entrepreneurs earn approximately 10% more than salaried workers with similar characteristics.

Entrepreneurs in this paper are the ones who declared themselves as self-employed in NLSY. This definition is similar to that used in previous studies of the returns to entrepreneurship. However, Hurst and Pugsley (2011) argue that these individuals are not necessarily the entrepreneurs that economic models and policy makers have in mind in that they have little desire to grow big or innovate in any observable way. For example, they may be small shopkeepers or restaurant owners. To the extent that even these shopkeepers and restaurant owners need to experiment with

their business ideas, and have the option to abandon in case of failure, the results of the paper go through. Arguably, there is even more experimentation going on in more innovative startups. This would in principle only make results stronger. A study that looked at data restricted to these more innovative startups would be an interesting avenue for further research.

The results of the paper are obtained for a particular cohort: individuals who were 14-22 years old in 1979. It is possible that the results will be different for different datasets or cohorts.[6] The National Longitudinal Survey of the Youth is following a new cohort whose respondents were 12-17 when first interviewed in 1997. It is too early to compute lifetime earnings for these individuals though.

However, Kerr, Nanda, and Rhoder-Kropf (2014) document a trend toward lower costs of experimentation in different industries. As argued here, the value of entrepreneurship arises from the real options that are available when experimenting with new ideas. A trend toward lower costs of experimentation is thus likely to make these options more valuable for more recent cohorts.

Another interesting topic for future research is that of serial entrepreneurs, that is, workers with more than one self-employment spell. Serial entrepreneurs represent approximately 4% of the sample in this paper, and thus there is not enough statistical power to study them.

References

Arabsheibani, G., D. De Meza, J. Maloney, and B. Pearson. 2000. And a vision appeared unto them of a great profit: Evidence of self-deception among the self-employed. *Economic Letters* 67:35–41.

Arrow, K. 1969. Classificatory notes on the production and diffusion of knowledge. *American Economic Review* 59:29–35.

Astebro, T. 2012. Returns to entrepreneurship. In D. Cummings (ed.), *The Oxford Handbook of Entrepreneurial Finance*. Oxford: Oxford University Press.

Astebro, T., H. Herz, R. Nanda, and R. Weber. 2014. Seeking roots of entrepreneurship: insights from behavioral economics. *Journal of Economic Perspectives* 28:49–70.

[6] See, for example, Astebro (2012) for a discussion of different datasets to study returns to entrepreneurship.

Bernardo, A., and I. Welch. 2001. On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy* pp. 301–330.

Blanchflower, D., and A. Oswald. 1992. Entrepreneurship, happiness, and supernormal returns: evidence from Britain and the U.S. Working Paper, National Bureau of Economic Research.

Cagetti, M., and M. C. De Nardi. 2006. Entrepreneurship, frictions, and wealth. *Journal of Political Economy* 114:835–70.

Campanale, C. 2010. Private equity returns in a model of entrepreneurship. *B.E. Journal of Maroeconomics, Contributions* 10:article 22.

Cooper, A., C. Woo, and W. Dunkelberg. 1988. Entrepreneurs' perceived chances for success. *Journal of Business Venturing* 3:97–108.

Fairlie, R. 2005. Entrepreneurship and earnings among young adults from disadvantaged families. *Small Business Economics* 25:223–36.

Gottlieb, J., R. Townsend, and T. Xu. 2016. Experimenting with entrepreneurship: the effect of job-protected leave. Working Paper, Dartmouth.

Hamilton, B. 2000. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy* 108:604–631.

Hintermaier, T., and T. Steinberger. 2005. Occupational choice and the private equity premium puzzle. *Journal of Economic Dynamics and Control* 29:1765–83.

Hombert, J., A. Schoar, D. Sraer, and D. Thesmar. 2015. Can unemployment insurance spur entrepreneurial activity? Working Paper, Berkeley-Haas.

Hurst, E., G. Li, and B. Pugsley. 2014. Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics* 96:19–33.

Hurst, E., and B. Pugsley. 2011. What do small business do? *Brookings Papers on Economic Activity* Fall 2011:73–118.

Kerr, W., R. Nanda, and M. Rhoder-Kropf. 2014. Entrepreneurship as experimentation. *Journal of Economic Perspectives* 28:25–48.

Kraus, A., and R. Litzenberger. 1976. Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085–1100.

Levine, R., and Y. Rubinstein. 2015. Smart and illicit: who becomes and entrepreneur and do they earn more? Working Paper, Berkeley-Haas.

Manso, G. 2011. Motivating innovation. *Journal of Finance* 66:1823–60.

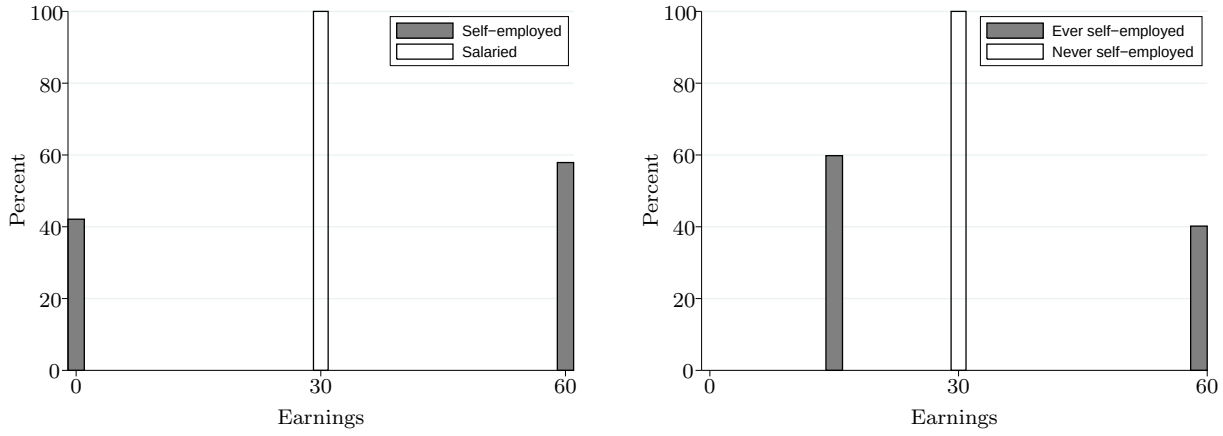
March, J. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2:71–87.

Moskowitz, T., and A. Vissing-Jorgensen. 2002. The returns to entrepreneurial investment: a private equity premium puzzle? *American Economic Review* 92:745–78.

Poschke, M. 2013. Who becomes an entrepreneur? Labor market prospects and occupational choice. *Journal of Economic Dynamics & Control* 37:693–710.

Schumpeter, J. 1934. *The theory of economic development.* Cambridge, MA: Harvard University Press.

Vereshchagina, G., and H. Hopenhayn. 2009. Risk taking by entrepreneurs. *American Economic Review* 99:1808–1830.



(a) Cross-sectional earnings distributions

(b) Lifetime earnings distributions

Figure 1. Comparing cross-sectional and lifetime earnings distributions by employment type
The figure presents cross-sectional and lifetime earnings distributions for self-employed and salaried workers. The parameters of the model used to generate these distributions are: $W = 30, R = 60, p = 0.4, \gamma = 0.05$. The figure shows that the cross-sectional self-employed earnings distribution is very different from the lifetime self-employed earnings distribution.

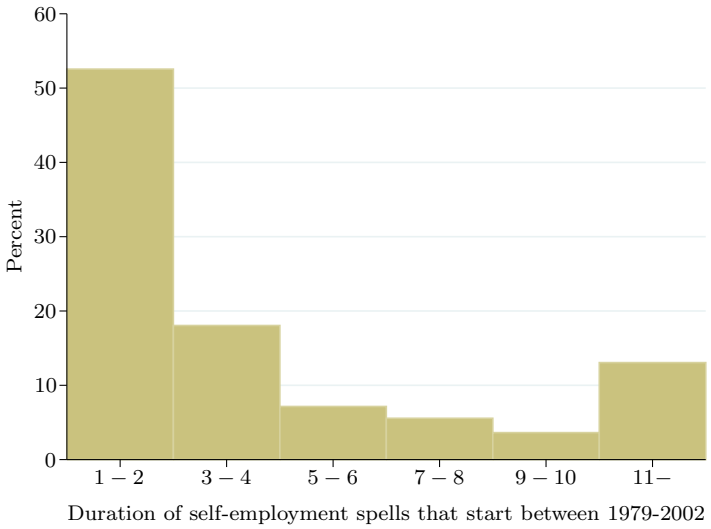


Figure 2. Duration of self-employment spells that start from 1979-2002

The figure presents the histogram of the distribution of the duration of self-employment spells that start from 1979-2002. Approximately 52% of self-employment spells last less than two years.

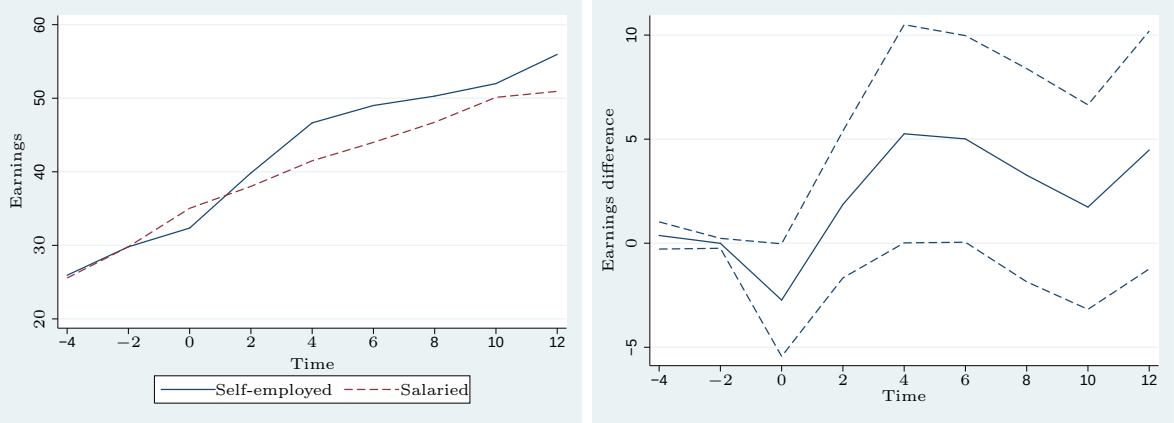


Figure 3. Earnings paths

Using propensity score methods, individuals who choose self-employment (treatment group) are matched with salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, prelabor market traits, industry, and year. The left graph shows mean earnings for the treatment group (solid line) and control group (dashed line) around the decision to become self-employed (time 0). The right graph shows the mean difference between the earnings of the treatment and control groups (solid line) around the decision to become self-employed (time 0) and the corresponding 95% gross confidence interval (dashed line).

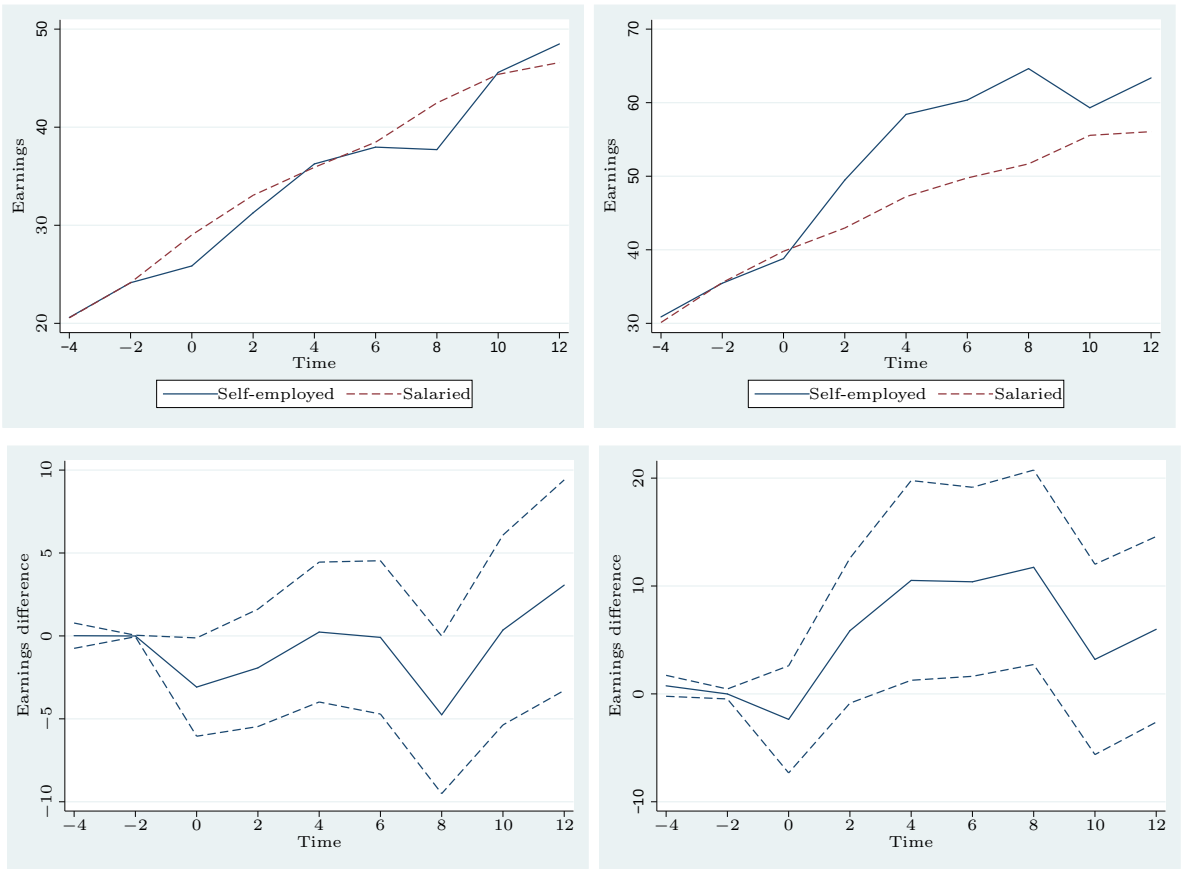


Figure 4. Earnings paths conditional on the duration of self-employment

Using propensity score methods, individuals who choose self-employment (treatment group) are matched with salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, prelabor market traits, industry, and year. In the upper-left figure, the solid line shows the mean earnings of individuals in the treatment group who stay entrepreneurs for less than 2 years, and the dashed line shows the mean earnings of the corresponding control group. In the upper-right figure, the solid line shows the mean earnings of individuals in the treatment group who stay entrepreneurs for more than 2 years and the dash line shows the mean earnings of the corresponding control group. The lower graphs show the corresponding mean differences between the earnings of the treatment and control groups (solid lines) around the decision to become self-employed (time 0) and the 95% gross confidence intervals (dashed lines).

	Salaried	Self-employed	Total
# of observations	52,303	5,470	57,773
Age	37.65	39.40	37.82
White	0.813	0.872	0.818
Female	0.510	0.385	0.498
Years of schooling	14.05	13.88	14.03
Weeks worked	46.36	46.65	46.39
Hours worked	1964.0	2101.6	1977.1
AFQT	49.36	51.01	49.52
Rotter locus of control	8.515	8.239	8.489
Rosenberg self-esteem	22.56	22.70	22.58
Mean annual earnings	47.04	52.25	47.57
Median annual earnings	37.36	31.13	36.87
SD annual earnings	45.75	73.12	48.25

Table 1. Summary statistics (pooled data)

The table presents summary statistics of 57,773 individual-year observations from 1980 to 2012 from the Bureau Labor of Statistics' National Longitudinal Survey of the Youth 1979 (NLSY79). The sample excludes earnings from people who are less than twenty-five years old, as well as earnings from people who do not work either as salaried or self-employed.

	Different models	
	(1)	(2)
	b/se	b/se
Residual self-employed earnings	-0.002***	-0.002***
	(0.000)	(0.000)
AFQT		-0.002**
		(0.001)
Rotter locus of control		0.019**
		(0.008)
Rosenberg self-esteem		-0.000
		(0.005)
Pseudo R-squared	0.010	0.012
Observations	4,785	4,785

Table 2. Effects of residual earnings on transition away from self-employment

The table presents coefficient estimates for a probit model estimating how residual earnings affect the probability of abandoning self-employment. The sample is restricted to self-employed workers over twenty-five years old. Residual earnings are obtained from an OLS regression that controls for demographics, educational attainment, work experience, industry, and prelabor market traits. Heteroscedasticity robust standard errors clustered at the year level are reported and *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Full sample	Subsamples	
		Currently salaried	Currently self-employed
	b/se	b/se	b/se
Previously self-employed	-0.286	5.161**	-3.907
	(1.505)	(2.181)	(5.747)
R-squared	0.651	0.703	0.608
Observations	36,832	33,581	3,251

Table 3. Effects of previous self-employment experience on earnings

The table presents coefficient estimates for an OLS model estimating how a previous completed self-employment spell affects annual earnings for individuals between 25 and 55 years old working full-time, year round. These regressions control for work experience, as well as year and individual fixed effects. Heteroscedasticity robust standard errors clustered at the year level are reported, and *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Never self-employed	Ever self-employed	Total
# of observations	3,863	1,552	5,415
White	0.795	0.843	0.809
Female	0.535	0.452	0.511
AFQT	47.08	49.62	47.81
Rotter locus of control	8.614	8.331	8.533
Rosenberg self-esteem	22.39	22.69	22.47
Mean annual earnings	40.46	42.77	41.12
Median annual earnings	36.11	34.67	35.59
SD annual earnings	34.48	41.79	36.38

Table 4. Summary statistics (lifetime)

The table presents summary statistics of 5,415 individuals from 1980 to 2012 from the Bureau Labor of Statistics' National Longitudinal Survey of the Youth 1979 (NLSY79). The sample excludes earnings from people who are less than twenty-five years old.

	Mean	
	Treatment	Control
Age	30.40	30.08
White	0.90	0.89
Female	0.49	0.51
Year of schooling	14.30	14.26
Work experience	10.13	9.86
AFQT	55.24	55.01
Rotter locus of control	8.26	8.28
Rosenberg self-esteem	22.77	22.69
Previous period earnings	29.80	29.81
Previous earnings growth	1.51	1.53

Table 5. Matching quality

The table reports mean values of treatment and control observable characteristics used in the matching procedure. Two-sided t-tests on the difference between mean values between treatment and control groups indicate no significant differences at the 95% confidence level for each variable.