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 are based on cross-sectional data and 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.
1 Introduction

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). I show that the estimates of mean and variance of entrepreneurial returns obtained by previous research are biased, as they are based on cross-sectional data and 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 as well as returns to entrepreneurship that are more attractive than established by previous research.

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, which is the only way to find out whether an idea is good. Alternatively, individuals may remain as salaried workers.

The model reveals that cross-sectional data analysis introduces 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 back to salaried workers or try new ventures.

These biases affect the estimates of mean and 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 tends to overstate 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 distribu-
tion of salaried earnings fails to account for the fact that the wage premium earned by salaried workers is 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 pre labor 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.

According to the model, self-employed workers experiment with new ideas when they leave the salaried workforce to become self-employed. 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 65% 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 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 and median lifetime earnings of the ever self-employed are higher than mean and median lifetime earnings of the never self-employed.

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, he is classified as ever self-employed, even though most of his earnings come from the time he was 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 as salaried worker. I find that on average, after becoming self-employed individuals earn more than their salaried counterparts. The median earnings difference between self-employed and salaried is not significantly different than zero.

Conditioning the analysis on the number of years as self-employed, I find that who attempted 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 to be self-employed. Individuals who stay as self-employed longer than two years experience 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 innovation. Schumpeter (1934) argues that entrepreneurship is essentially the experimentation with “new combinations” of existing resources. Arrow (1969) associates innovation with the production of knowledge and proposes the use of Bayesian decision models to study innovation. Bandit problems are Bayesian decision models that allow for knowledge acquisition through experimentation. Weitzman (1979) applies a simple bandit problem to study the innovation process. March (1991) uses the terms exploration and exploitation to describe the fundamental tension that arises in learning through experimentation. Manso (2011) shows that tolerance for failure and reward for long-term success are optimal to motivate exploration. Kerr, Nanda, and Rhoder-Kropf (2014) survey the literature on entrepreneurship as experimentation.

In this paper, I focus on showing that previous estimates of the returns to entrepreneurship are biased because they are based on cross-sectional data. Other papers try to resolve the private equity premium puzzle by providing explanations for why entrepreneurs might accept to work for less. For example, entrepreneurs may enjoy non-pecuniary benefits (Blanchflower and Oswald 1992), have a preference for skewness (Kraus and Litzenberger 1976), or be overconfident (Cooper, Woo, and Dunkelberg 1988, Bernardo and Welch 2001).

Levine and Rubinstein (2012) argue that self-employment is not a good measure of entrepreneurship. They show that self-employed workers who incorporate their business earn sub-
stantially more than salaried worker and argue that only self-employed 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. Hintermaier and Steinberger (2005), Vereshchagina and Hopenhayn (2009), and Campanale (2010) develop dynamic models of occupational choice in which workers can choose to become entrepreneurs and learn about their entrepreneurial skills. However, because the models do not feature overlapping generations they cannot be used to compare the cross-sectional and lifetime distributions of earnings. Moreover, these papers do not contain any empirical analysis.

The paper is organized as follows. Section 2 contains the basic model of entrepreneurship as experimentation. Section 3 extends the model to allow for a premium or discount to past entrepreneurial experience. Section 4 uses the NLSY79 data to establish patterns that are consistent with entrepreneurship as experimentation and compares lifetime earnings of entrepreneurs and salaried workers. Section 5 concludes.

2 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 $\gamma$ 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 finds the idea is of 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 out about the quality of a new idea is by trying it out as self-employed.

To capture the exploratory nature of self-employment, I assume that $R > W$ and $pR < W$. If successful, 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 as 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 \( R > W \). If it turns out that their idea is of low quality, it yields 0, and they will abandon it and return to the salaried workforce. The expected per period earnings \( V_{\text{semp}} \) of such strategy are:

\[
V_{\text{semp}} = pR + (1 - p)\frac{W}{2}
\]  

(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 salaried workers if and only if \( V_{\text{semp}} \geq V_{\text{sal}} \), which is equivalent to

\[
pR \geq \frac{(1 + p)}{2} W
\]  

(2)

Otherwise, agents earn more as salaried workers.

The above comparison only takes into account monetary payoffs of workers. To ensure that agents with ideas try them out, we assume that self-employed workers enjoy private benefits \( \beta \) that are high enough to make entrepreneurship payoff. 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 \( \theta_{\text{sal}} \), \( \theta_{\text{semp,s}} \), and \( \theta_{\text{semp,f}} \) 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:

\[
\theta_{\text{sal}} = (1 - \gamma) + \frac{\gamma(1 - p)}{2}
\]

\[
\theta_{\text{semp,f}} = \frac{\gamma(1 - p)}{2}
\]

\[
\theta_{\text{semp,s}} = \gamma p
\]

\[1\]In the model of this section, this condition holds if \( \beta \geq W - 2pR/(1 + p) \).
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 back to salaried workers 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: $W = 30, R = 60, p = 0.35, \gamma = 0.05$. These distributions illustrate the survivorship and experimentation biases. Due to the 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 the 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 back to the salaried workforce earning a lifetime mean payoff that is not only greater than zero, but only slightly below $W$.

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:

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

which is greater than $V_{semp}$ if

$$pR + \frac{(1 - p)W}{2} \geq \frac{2p}{1 + p} R \iff pR \geq \frac{(1 + p)}{2} W.$$
The survivorship bias leads to an overstatement of the true lifetime mean of self-employed earnings. The 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 the 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 self-employed earnings with the lifetime standard deviation of self-employed earnings.

**Proposition 2** There exist $\lambda > 1$ such that the cross-sectional standard deviation of self-employed earnings overstates the lifetime standard deviation of self-employed earnings if and only if $\lambda = \frac{V_{s Kemp}}{V_{s Sal}} < \lambda$.

**Proof** The cross-sectional variance of self-employed earnings is:

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

(3)

The lifetime variance of self-employed earnings is:

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

(4)

which is increasing in $\lambda$ and lower than (3) if $\lambda = 1$ (or $V_{s Kemp} = V_{s Sal}$).

The experimentation bias amplifies entrepreneurial failures, while the survivorship bias overweights successful entrepreneurs, both in principle contributing to an overstatement of the variance of self-employed earnings. It is only for extreme cases in which the survivorship bias makes the cross-sectional mean of self-employed earnings much higher than the lifetime mean of self-employed earnings that the cross-sectional variance may overstate the lifetime variance of self-employed earnings.

### 3 Prior Entrepreneurial Failure

In this section we consider the same model as in the previous section, except that after experiencing an entrepreneurial failure, salaried workers earn $\kappa_1 W$. 

7
Self-employed lifetime mean earnings are:

\[ V_{\text{semp}} = pR + (1 - p)\frac{\kappa_1 W}{2} \]

while salaried workers lifetime mean earnings are \( V_{\text{sal}} = W \).

The cross-sectional distribution needs to account for fractions of successful self-employed, failed self-employed, as well as salaried workers with and without a previous entrepreneurial failure. The cross-sectional distribution is given by:

\[
\begin{align*}
\theta_{\text{semp},s} &= \gamma p \\
\theta_{\text{semp},f} &= \frac{\gamma(1 - p)}{2} \\
\theta_{\text{sal}(1)} &= \frac{\gamma(1 - p)}{2} \\
\theta_{\text{sal}(0)} &= (1 - \gamma)
\end{align*}
\]

**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:

\[
\frac{\theta_{\text{sal}(0)}W + \theta_{\text{sal}(1)}(\kappa_1 W)}{\theta_{\text{sal}(0)} + \theta_{\text{sal}(1)}} = \frac{2(1 - \gamma) + \gamma(1 - p)\kappa_1}{2(1 - \gamma) + \gamma(1 - p)} W
\]

which is greater than \( W \) iff \( \kappa_1 > 1 \). ■

If previous entrepreneurial failures improve salaried worker earnings \( (\kappa_1 > 1) \), 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 the worker has formerly been self-employed or not. Due to such attribution bias, the cross-sectional mean earnings of salaried workers overstates the lifetime mean of salaried workers earnings.

**Proposition 4** There exists \( \bar{\kappa} \) such that for \( \kappa_1 > \bar{\kappa} \) 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:

\[ V_{cs} = \frac{\theta_{s\text{emp},s}}{\theta_{s\text{emp},s} + \theta_{s\text{emp},f}} R = \frac{2p}{1 + p} R, \]

which is independent of \( \kappa_1 \). However, \( V_{s\text{emp}} \) is increasing in \( \kappa_1 \) and goes to infinity as \( \kappa_1 \) goes to infinity.

The cross-sectional distribution of self-employed earnings does not take into account the salaried wage premium \( \kappa_1 \) that only accrues to the worker if he has previous experience as self-employed. For high enough \( \kappa_1 \), this attribution bias becomes dominant and the cross-sectional mean of self-employed earnings understates the lifetime mean of self-employed earnings.

4 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.

4.1 Data

The NLSY79 is a survey of 12,686 individuals who were 15-22 years old when they were first surveyed in 1979. The same individuals were then surveyed annually through 1994 and biennially thereafter. The last survey year in my sample is 2010.

The NLSY79 cohort is comprised of three subsamples. For this study, I drop the military and representative minorities subsamples, and restrict the analysis to the nationally representative subsample of 6,111 individuals. I dropped from the sample observations in which an individual has missed the survey. I also dropped from the sample all observations corresponding to individuals who were never a worker during the period 1979-2010. The final sample contains observations on 5,409 individuals.

From the NLSY79, I obtain information on demographics, educational attainment, labor market outcomes, and pre labor 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, and weeks worked.
I consider an individual a worker (either self-employed or salaried) during a given year if the individual reported at least 300 hours of work during the year.

Pre labor market traits include three different measures. To measure cognitive ability, I use AFQT score (Armed Forces Qualifications Test 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 oneself. The values range from six to 30, with higher values 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 16, where higher values signify less internal control and more external control.

All earnings variables are adjusted for inflation using the Consumer Price Index (CPI). Earnings are expressed in 2010 dollars.

4.2 Cross-sectional Earnings

In this subsection, I take individual-year observations as the basic unit of analysis, to 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 of self-employed workers are only slightly higher than those of salaried workers. However, median annual earnings of self-employed workers are lower than those of salaried workers. Moreover, the standard deviation of self-employed
earnings is substantially higher than that of salaried workers earnings. These are in line with previous studies which conclude that the median self-employed individual earn less and bear significantly more risk than salaried workers. Even though mean self-employed earnings are slightly 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.

[Table 1 about here.]

The findings in Table 1 are in line with previous results in the literature. Mean self-employed earnings are higher than mean salaried workers earnings, while median self-employed earnings are lower than median salaried earnings. Entrepreneurship seems risky as the standard deviation of self-employed earnings is higher than the standard deviation of salaried earnings.

4.3 Lifetime Earnings

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

According to the model in Section 2, self-employed workers are engaging in experimentation when they choose self-employment. As such, the average self-employment spell 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 as self-employed.

Figure 2 shows a histogram with the duration of self-employment spells that start between 1979 and 2000. Approximately 65% of self-employment spells last less than 2 years. This is consistent with the view that 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 OLS
regression shown in Table ??, which explains earnings based on employment type, demographics, educational attainment, and pre labor market traits. According to Table ?? lower residual earnings while self-employed are associated with a higher probability of abandoning self-employment.

[Table 2 about here.]

The basic model of Section ?? also predicts that self-employed workers are not penalized if they decide to go back to the salaried workforce. Table ?? shows that indeed workers with previous self-employment experience earn a premium when compared to similar workers without self-employment experience. This premium is higher for self-employed workers who have previously completed a self-employment spell. These findings are consistent with the extension of the basic model considered in Section ?? with a premium for previous self-employment experience ($\kappa_1 > 1$).

[Table 3 about here.]

Using NLSY79, we can compute lifetime mean earnings for different employment types. Table ?? provides summary statistics and is the analogous of Table ?? 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. In contrast to Table ?? which is based on cross-sectional data, Table ?? shows that median (and mean) lifetime earnings of those who have ever been self-employed are higher than median (and mean) lifetime earnings of those who have never been self-employed. Moreover, the standard deviation of lifetime earnings of workers who have been self-employed is not substantially higher than the standard deviation of lifetime earnings of workers who have never been self-employed.

[Table 4 about here.]

The results in Table ?? call into question previous findings which claim that entrepreneurs earn less and bear significantly more risk than salaried workers. Once lifetime earnings are taken into account, returns of entrepreneurs are higher than those of salaried workers, while entrepreneurs bear only a little more risk than salaried workers.
4.4 Propensity Score Matching

The comparison in the previous section was between never self-employed and ever self-employed individuals. Such 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 with someone who looks just like this individual in terms of observed characteristics but decides to remain as salaried worker.

To achieve such goal, I run a propensity score matching analysis. Individuals are matched based on path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry, occupation, and year. I use the nearest neighbor matching algorithm without replacement.

Figure 3 compares outcomes between treatment and control groups. As shown in the figure, after becoming self-employed, individuals earn on average approximately $3,000 more per year than similar individuals who decided to remain as salaried workers. This gain sustained throughout their lives.

[Figure 3 about here.]

Figure 4 shows that the difference is statistically significant in most years. When taken as the average of post self-employed years, the difference is also statistically significant.

[Figure 4 about here.]

Figure 5 shows that the median self-employed earnings are indistinguishable between the two groups. This should be contrasted with the results in the previous literature, which had found that the median self-employed earnings were significantly lower than the median salaried earnings.

[Figure 5 about here.]

Figure 6 compares treatment and control groups conditional on the duration of entrepreneur-ship span. Panel A shows the results for individuals whose self-employment spans last less than 2
years, while panel B shows the results for individuals whose self-employment spans last more than 2 years. Individuals who attempt to be entrepreneurs but abandon entrepreneurship in less than two years, are not punished, achieving approximately 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.

This figure 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.

[Figure 6 about here.]

Figure 7 shows the difference between mean earnings of individuals in the treatment and control groups around the decision to become entrepreneurs (time 0). In the left figure, the solid line shows mean difference earnings between individuals in the treatment group who stay as entrepreneurs for less than two years and their pairs. The dashed line represents 95%-confidence intervals. The right figure is analogous for individuals in the treatment group who stay as entrepreneurs for more than two years.

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, these individuals do not perform significantly different from their matched pairs who were never an entrepreneur. Entrepreneurs who decide to stay longer than two years, do not suffer at time 0 and perform significantly better than their matched pairs afterwards.

[Figure 7 about here.]

Figure 8 shows the distribution of earnings growth for treatment and control group around the decision to become entrepreneur. The left panel shows the distribution of relative earnings growth, while the right panel shows absolute earnings growth. In both cases, the gamble taken by entrepreneurs has higher mean without significant extra risk.

[Figure 8 about here.]
5 Conclusion

References


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: $W = 30$, $R = 60$, $p = 0.35$, $\gamma = 0.05$. The figures show 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 between 1979-2000

The figure presents the histogram of the distribution of the duration of self-employment spells that start between 1979-2000. Approximately 65% of self-employment spells last less than two years.
Figure 3: Using propensity score methods, individuals who choose to become self-employed (treatment group) are matched with individuals who stay as salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry occupation, and year. The figure shows mean earnings for treatment group (solid line) and control group (dashed line) around the decision to become self-employed (time 0).
Figure 4: Using propensity score methods, individuals who choose to become self-employed (treatment group) are matched with individuals who stay as salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry occupation, and year. The figure shows the mean difference between earnings of treatment and control groups (solid line) around the decision to become self-employed (time 0) and the corresponding 95%-confidence interval.
Figure 5: Using propensity score methods, individuals who choose to become self-employed (treatment group) are matched with individuals who stay as salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry occupation, and year. The figure shows the median difference between earnings of treatment and control groups around the decision to become self-employed (time 0).
Figure 6: Using propensity score methods, individuals who choose to become self-employed (treatment group) are matched with individuals who stay as salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry occupation, and year. In the left figure, the solid line shows mean earnings of individuals in the treatment group who stay as entrepreneurs for less than 2 years and the dash line shows mean earnings of the corresponding control group. In the right figure, the solid line shows mean earnings of individuals in the treatment group who stay as entrepreneurs for more than 2 years and the dash line shows mean earnings of the corresponding control group.
Figure 7: Using propensity score methods, individuals who choose to become self-employed (treatment group) are matched with individuals who stay as salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry occupation, and year. In both figures, the solid line shows mean difference between earnings of individuals in the treatment and control groups and the dashed lines represent the corresponding 95%-confidence interval. The left (right) figure is conditional on individuals in the treatment group staying as entrepreneurs for less (more) than two years.
Figure 8: Using propensity score methods, individuals who choose to become self-employed (treatment group) are matched with individuals who stay as salaried workers (control group) based on earnings path before treatment, work experience, demographics, educational attainment, pre-labor market traits, industry occupation, and year. The left figure shows histograms of relative earnings growth around the decision to become entrepreneur (time 0) for treatment and control groups. The right figure shows histograms of absolute earnings growth around the decision to become entrepreneur (time 0) for treatment and control groups.
<table>
<thead>
<tr>
<th></th>
<th>Salaried</th>
<th>Self-Employed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Observations</td>
<td>61166</td>
<td>1941</td>
<td>63107</td>
</tr>
<tr>
<td>Age</td>
<td>34.14</td>
<td>32.15</td>
<td>34.08</td>
</tr>
<tr>
<td>White</td>
<td>0.822</td>
<td>0.894</td>
<td>0.824</td>
</tr>
<tr>
<td>Female</td>
<td>0.490</td>
<td>0.344</td>
<td>0.486</td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>13.98</td>
<td>13.71</td>
<td>13.97</td>
</tr>
<tr>
<td>Weeks Worked</td>
<td>46.94</td>
<td>47.08</td>
<td>46.94</td>
</tr>
<tr>
<td>Hours Worked</td>
<td>1982.4</td>
<td>2184.9</td>
<td>1988.6</td>
</tr>
<tr>
<td>AFQT</td>
<td>49.67</td>
<td>52.08</td>
<td>49.74</td>
</tr>
<tr>
<td>Rotter Locus of Control</td>
<td>8.506</td>
<td>8.158</td>
<td>8.496</td>
</tr>
<tr>
<td>Rosenberg Self-Esteem</td>
<td>22.58</td>
<td>22.90</td>
<td>22.59</td>
</tr>
<tr>
<td>Mean Annual Earnings</td>
<td>38.25</td>
<td>40.55</td>
<td>38.32</td>
</tr>
<tr>
<td>Median Annual Earnings</td>
<td>29.76</td>
<td>26.38</td>
<td>29.65</td>
</tr>
<tr>
<td>SD Annual Earnings</td>
<td>38.42</td>
<td>47.47</td>
<td>38.70</td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics (Pooled Data)
This table contains summary statistics for the sample of 70,296 individual-year observations from 1979 to 2010.
<table>
<thead>
<tr>
<th>Different Models</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td></td>
</tr>
<tr>
<td>b/se</td>
<td>b/se</td>
<td></td>
</tr>
<tr>
<td>Residual Self-Employed Earnings</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>AFQT</td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>Rotter Locus of Control</td>
<td>0.042***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Rosenberg Self-Esteem</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Effects of residual earnings on transition away from self-employment
This table presents coefficient estimates for a probit model estimating how residual earnings affect the probability of abandoning self-employment. These regressions control for demographics, educational attainment, work experience, occupation, industry, and pre labor market traits. The symbols *, **, *** represent significance at the 10%, 5%, and 1% level.
Table 3: Effects of previous self-employment experience on earnings

<table>
<thead>
<tr>
<th></th>
<th>Different Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Previously Self-Employed</td>
<td>1.577***</td>
</tr>
<tr>
<td></td>
<td>(0.449)</td>
</tr>
<tr>
<td>(Previously Self-Employed)*(Self-Employed)</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td>(2.317)</td>
</tr>
</tbody>
</table>

This table presents coefficient estimates for an OLS model estimating how previous self-employment experience affects earnings. These regressions control for demographics, educational attainment, work experience, occupation, industry, and pre labor market traits. The symbols *, **, *** represent significance at the 10%, 5%, and 1% level.
<table>
<thead>
<tr>
<th></th>
<th>Salaried</th>
<th>Self-Employed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Observations</td>
<td>4450</td>
<td>926</td>
<td>5376</td>
</tr>
<tr>
<td>White</td>
<td>0.797</td>
<td>0.869</td>
<td>0.810</td>
</tr>
<tr>
<td>Female</td>
<td>0.533</td>
<td>0.392</td>
<td>0.509</td>
</tr>
<tr>
<td>AFQT</td>
<td>47.20</td>
<td>51.40</td>
<td>47.92</td>
</tr>
<tr>
<td>Rotter Locus of Control</td>
<td>8.588</td>
<td>8.257</td>
<td>8.531</td>
</tr>
<tr>
<td>Rosenberg Self-Esteem</td>
<td>22.41</td>
<td>22.79</td>
<td>22.47</td>
</tr>
<tr>
<td>Mean Annual Earnings</td>
<td>38.69</td>
<td>45.52</td>
<td>39.87</td>
</tr>
<tr>
<td>Median Annual Earnings</td>
<td>35.64</td>
<td>39.82</td>
<td>36.36</td>
</tr>
<tr>
<td>SD Annual Earnings</td>
<td>29.98</td>
<td>36.43</td>
<td>31.09</td>
</tr>
</tbody>
</table>

Table 4: Summary Statistics (Lifetime)
This table contains summary statistics for the sample of 5,409 individuals from 1979 to 2010.