IFN Working Paper No. 1050, 2014

Does College Education Reduce Small Business Failure?

Andrea Asoni and Tino Sanandaji

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Does college education reduce small business failure?*

Andrea Asoni† Tino Sanandaji ‡

November 30, 2014

Abstract

We estimate the effect of college education on business survival using the NLSY79. The endogeneity of both education and business ownership is accounted for by a competing risks duration model augmented with a college selection equation. Contrary to the previous literature, we find no effect of college education on business failure. College however significantly increases "employment survival." Unlike college, cognitive skills have a positive impact on employment survival for both the self-employed and employees. The results suggest that college affects the self-employed and salaried employees in different ways, for example generating skills more useful in employment than self-employment.

Key words: Selection, Entrepreneurship, College Education, Intelligence, Self-confidence

JEL *classification:* C41, J24, L26

*We thank Luigi Zingales, Erik G. Hurst, Robert J. LaLonde, Susanne M. Schennach, Steven N. Kaplan, Adair Morse, Luigi Guiso, Margarita Tsoutsoura, James J. Heckman, the participants to many seminars at the University of Chicago and the participants to the Brownbag Seminar at Boston Fed for useful comments. Jacob Bergmann Larsen provided excellent research assistance. All remaining errors are ours. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

†Corresponding author: asoni@uchicago.edu. Charles River Associates, Washington DC & IFN – Research Institute for Industrial Economics. Stockholm, Sweden. The views presented here are my own and do not necessarily reflect those of CRA or any CRA employee

‡IFN – Research Institute for Industrial Economics. Stockholm, Sweden.

1 Introduction

A college degree is generally found to be valuable for salaried employees even after accounting for skills (for example, Angrist and Krueger 1991, Card 1995, Card 1999). The earning premium associated with a college degree may be due to both human capital formation and the signaling role of education used by employers to screen for ability (Spence 1974, Heckman et al. 2006, Arcidiacono et al. 2010). Less is known about the effect of a college degree on the labour market outcomes of the self-employed.

Business owners with college degrees are generally found to be more successful, but selection bias and skill levels are rarely accounted for (Van Praag et al., 2009). In particular, the endogeneity caused by the relationship between innate characteristics, level of education attained and decision to become a self-employed has not been properly taken into account. In this paper we propose an empirical approach that controls for both the effect of innate characteristics on educational choices, and the effect of innate characteristics and education attained on occupational choices. Our identification strategy relies on a combination of instrumental variables for college attainment and controls for cognitive skills. Moreover, we employ a competing risks duration model with unobserved heterogeneity to control for selection into entrepreneurship.

To our knowledge this is the first time that this approach has been used to study the effect of college education on business success.

We estimate our model on a rich dataset (the National Longitudinal Survey of Youth - 1979; NLYS79 from now on) and we find that college does not have an effect on business success after controlling for the endogeneity of education and occupational choices. In this paper we define "business success" as business survival.

We additionally show that the aforementioned result is true only for self-employed and not for salaried employees. College education does have a positive effect on "employnment survival" for salaried workers. This suggests that self-employed and salaried workers rely on partially different sets of skills to be successful in their work. We investigate separately men and women and, while we find that they differ along some important dimensions in firm ownership and dynamics, our results are consistent across men and women.

The next section describes the previous literature. The third section discusses the problems associated with correctly estimating the effect of college education on business survival; we then describe our identification strategy and the model used in this paper. The fourth section discusses the data, the fifth our results and the sixth concludes.

2 Human Capital and Business Dynamics in the Literature

Numerous studies have investigated the importance of college education for the self-employed. Two recent meta studies of this literature (Van der Sluis et al 2008, Unger et al 2011) point to a positive association between college education and various measures of business success. In this study we focus on one measure of business success, namely business survival. Focusing on this outcome measure allows us to account for the dual endogeneity of education and business ownership within the framework of our model. The effect of college on business survival among the self-employed is compared with the effect of college on the risk of employment termination for salaried employees.

Van der Sluis et al. (2008) report that most of about thirty surveyed studies find that education improves survival. Bruderl et al. (1992), Gimeno et al. (1997), Boden and Nucci (2000), Lin (2000), Millan et al. (2010), Rauch et al. (2011) and Ganotakis (2012) are examples of studies finding a positive link between education and business survival. By contrast Davidsson and Honig (2003), Van Praag (2003) and Blanchflower and Meyer (1994) find that a college degree is associated with higher failure rates. Few studies account for unobserved skills and selection. Moreover, studies have often not distin-

guished between exit from self-employment into employment and exit into non-employment. Transition from self-employment to employment may represent an upward or lateral career move. Only transition into non-employment can be plausibly defined as business failure, though this too may in some instances reflect voluntarily exit. We focus on transition to non-employment to measure business survival.

3 An Empirical Model of Occupational and Educational Choices

3.1 Challenges to the Estimation of the Effect of College Education on Firm Dynamics

Identifying the effect of education on business survival requires addressing several sources of bias. First, the strong link between ability and education causes an upward bias on the effect of schooling (Becker, 1964). Second, even when the analysis includes measures of ability, there are still unobserved characteristics that simultaneously affect educational choices and labor market outcomes. For instance more entrepreneurial individuals can substitute college education with industry-experience (for example, dropping out of college to

start a business). This will translate into a negative bias for the measured effect of education as the most talented entrepreneurs are less likely to complete their studies.

Third, even in a scenario in which college education is entirely exogenous - and the issues discussed above are not relevant - researchers still need to take into account the fact that college education itself affects the decision to become self-employed. College education in fact might change the distribution of returns for employees and self-employed differently. For example, if market returns from a college degree are higher for employees than the self-employed only the college educated with most business talent will create firms. This will positively bias the measured effect of college education on entrepreneurial success. If the reverse is true, as it has recently been suggested (for example, Hartog et al., 2010; van Praag et al., 2013), the bias will be negative.

The fourth and final issue to be considered is what Heckman and Singer (1985) call "negative duration dependence" bias. Those who are most likely to transition from one state (self-employment) to another (non-employment) on average also tend to experience such transition sooner. In other words, the composition of the population of self-employed, in terms of unobserved heterogeneity, changes over time as people move from self-employment to employ-

ment or non-employment. For instance individuals with lower entrepreneurial talents are more likely to transition out of self-employment and they will do so sooner rather than later. Ceteris paribus this implies that the average entrepreneurial talent of the self-employed is increasing over time. If, for example, those entrepreneurs are also the most educated entrepreneurs this will bias the measured effect of education on survival.

3.2 Addressing the Challenges

In order to cope with these issues we rely on an multi-pronged empirical strategy. The NLSY79 includes accurate measures of cognitive skill, reducing an important element of unobserved ability bias. Most importantly, we apply a competing risks duration model with unobserved heterogeneity as developed by LaLonde and Ham (1996), and modified by Eberwein et al. (1997). The key to identification is analysing the decisions to create a firm and the subsequent decision to continue or terminate the firm simultaneously. This way we can use information about the individual elicited through the entry decision (starting a firm) to better account for the probability of exit (moving to non-employment). This model addresses the relationship between education and selection into entrepreneurship, and the relationship between unobserved entrepreneurial talents, selection into self-employment and self-employment

outcomes. It does so by imposing a flexible structure on the unobserved heterogeneity and estimating it jointly with the rest of the parameters of the model. Below we describe the details of this model.

To address the endogeneity of education to both ability and entrepreneurial talents we augment this model with a selection equation into college. This approach is analogous to Eberwein et al. (1997). Identification comes in this case from two instruments for educational attainment. The first instrument is distance from college interacted with parental education. Distance from college at the time of high school graduation creates variation in the cost of obtaining a degree. One problem is that college educated parents tend to live closer to universities. The solution proposed by Card (1995) is interacting distance from college with parental education. The instrument is a dummy variable that is equal to one if parents lack a college degree and the individual lives in a county where a college is located, and zero otherwise.

The second instrument is local unemployment rate at the time when the individual graduated from high school. During economic booms, high school graduates are more likely to join the labor force and not continue their studies. When labor market demand is weaker attending colleges becomes more attractive. The state of the labor market rate the year in which one graduates

can be assumed to be largely random with respect to ones educational choices. Our instrument is defined as the deviation in the year of completion of high school degree from the average local unemployment rate over time.

3.3 Modeling Occupational Choice

3.3.1 A Competing Risks Model

For expositional ease let us start by assuming that education is exogenous and focus on occupational choice. We rely on a competing risks, multinomial choice model in a dynamic setting based on Ham and LaLonde (1996). At any given time each individual is in one of three mutually exclusive states: self-employment (s), employment (e) or non-employment (n). At the end of each period one of three things will happen: he stays in his current state, or he moves to one of the other two states.

Consider someone who is employed at the beginning of the time period covered by the data. At the baseline, t_B, this person is employed. After t_e periods he starts a firm and manages it for t_b periods. The business spell is followed by a non-employment spell of length t_n that is right censored because we have reached the end of the time period covered by the data. We are interested only in the probability of moving out of a business spell, but we will

look at the entire employment history for an unbiased estimate.

The transition probabilities across states depend on demographic as well as macroeconomic variables and unobserved characteristics. They can be written as follows: in any given period the probability $\lambda_{i,kj}$ that individual i leaves state k and enters state j is

$$\lambda_{i,kj}(t|\theta_{i,kj}) = \frac{\exp(y_{i,kj}(t))}{1 + \sum_{s \neq k} \exp(y_{i,ks}(t))}$$

with

$$y_{i,kj}(t) = \beta_{kj} X_i(t) + \gamma_{kj} CO_i + h_{kj}(t) + \theta_{i,kj}$$

where $X_i(t)$ is a vector of control variables containing race, age, age squared, marital status, number of kids and local unemployment rate, measured as log deviation from local average; industry dummies are also included to capture differences in the industrial structure of different industries that might influence transition rates. CO_i is a dummy variable indicating whether the individual has a (4-years) college degree. $h_{kj}(t)$ is a function of duration; in particular $h_{kj}(t) = \delta_{1,kj} \log(t) + \delta_{2,kj} \log^2(t)$. We also consider an alternative specification that includes cognitive skills (measured by the AFQT) as well as a measure of non-cognitive ability (Locus of Control) among the controls. We

hope that this will lead to better identification of the college effect.

The unobserved heterogeneity is captured by the scalar random variable $\theta_{i,kj}$. These unobserved factors are assumed to be fixed across spells of the same type and potentially different across spells of different types. There are six different $\lambda_{i,kj}(\cdot|\theta_{i,kj})$ since there are 3 states: $kj = es$, en, se, sn, ne and ns. However we impose that the unobserved factor that influences, for example, exit rates from business spell is the same irrespectively of the exit route taken. Same thing is true for employment and non-employment spells. In other words we impose $\theta_{es} = \theta_{en} = \theta_e$, $\theta_{se} = \theta_{sn} = \theta_s$, and $\theta_{ne} = \theta_{ns} = \theta_n$.

3.3.2 Unobserved Heterogeneity and Occupational Choice

Unobserved heterogeneity is important in this model for several reasons. First, it helps us correct for the existence of negative duration dependence bias. Let us explain how: for expositional ease assume that there is no relationship between length of self-employment spell (duration) and probability of failing - that is start-ups and well-established firms have the same probability of going bankrupt ceteris paribus. Furthermore, assume that there are only two types of business owners, skilled and unskilled. Unskilled business owners are more likely to fail and move to non-employment (or employment). As the more unskilled business owners leave the self-employed group only the best among

them are left managing firms. This creates in the data a negative relationship between duration and exit probability where none exists.

In order to deal with this problem we follow a standard approach in the labor literature and specify a distribution for the unobserved heterogeneity terms. The parameters of such distribution need to be estimated along with the rest of the coefficients. By specifying a distribution of the unobserved terms we can write down the average likelihood function and then estimate the *average* survival function, correcting for the negative duration dependence bias.

Second, the unobserved heterogeneity helps us deal with selection into self-employment *induced by education*. Even if college education was entirely exogenous and not related to underlying unobserved factors, there could still exists a systematic relationship between college education and the unobserved characteristics of those in the self-employment group. Consider the following example. Each individual has only two dimensions: business talent and education. Business talent is not observed by the econometrician. Education is randomized in the population and, by construction, not correlated with business talent. However, more educated people earn higher wages in the "employed" sector and this makes them less likely to start a business. As a consequences only the most talented among the highly educated will self-select

into entrepreneurship. This creates a positive correlation among businessmen between education and business talent.[1] Another way to state it is that education changes the distribution of unobservables in the subpopulation of business owners.

One way to deal with this problem is to allow the unobserved characteristics to be correlated across spells of different types. In particular, we allow the unobserved term in any spell preceding a business period to be correlated with the unobserved factor in the business spell. This way by analyzing the process of selection into self-employment we learn what type of individuals education makes more likely to enter self-employment, and we can use this information when estimating the effect of education itself: once we have learned how education changes the distribution of unobservables in the self-employed population we can use this information to average out the effect of the unobserved terms and estimate the average effect of education on business spells.

The third issue where unobserved factors play a role is with left-censored spells. The model we have described so far is conditional on the initial distribution of individuals across states. Instead of explicitly modelling such distribution we allow the heterogeneity term of the first (left-censored) spell

[1]Note that one could come up with an example with a negative correlation between education and business ownership. The model is not predicated on the relationship being positive. In fact the sign of the relationship will be estimated together with the rest of the parameters of the model.

to be different from the heterogeneity term of the corresponding spell type in the rest of the employment history (as suggested in Ham and LaLonde, 1996). Consider an individual who at the baseline is employed. For this individual the first, left censored spell is an "employment spell". For all the other employment spells that this person will experience during his working life, the heterogeneity term will be θ_e. For this first left censored spell the heterogeneity term instead is θ_{el}, with $\theta_e \neq \theta_{el}$ (where l stands for "left censored"). We define analogously θ_{sl} and θ_{nl}.

We assume that the distribution of the unobserved terms has a finite number of points. In particular for $k = e,\ s,\ n$:

$$\begin{aligned} \theta_{i,k} &= c_k\theta_1^* \qquad \text{with probability } P_k \\ &= c_k\theta_2^* \qquad \text{with probability } 1 - P_k \end{aligned}$$

where c_k, θ_1^*, θ_2^* and the probability P_k are estimated along with the rest of the parameters. For identification reasons c_n is normalized to 1. The loading factors c_k describe the relationship between the unobserved factor in the non-employment spells with the others (and hence of any spell with all the others).

If they are positive then there is a positive relationship between unobserved factors, and viceversa if negative.

3.3.3 The Likelihood Function

We can write the likelihood of any employment history using the aforementioned probabilities. In general, the probability of a spell of type k of length t_k that ended with transition in state j as opposed to state m is written as:

$$f_{kj}(t_k|\theta_{kj},\theta_{km}) = \lambda_{kj}(t_k|\theta_{kj}) \prod_{s=1}^{t_k-1} (1 - \lambda_{kj}(s|\theta_{kj}) - \lambda_{km}(s|\theta_{km}))$$

Analogously the contribution of a right censored type-k spell can be described as:

$$S_k(t_k|\theta_{kj},\theta_{km}) = \prod_{s=1}^{t_k-1} (1 - \lambda_{kj}(s|\theta_{kj}) - \lambda_{km}(s|\theta_{km}))$$

To write the contribution to the likelihood function of the employment history described at the beginning of this section we need to integrate the probability of each spell over the distribution of the unobserved characteristics:

$$\int f_{eb}(t_e|\theta_{el}) f_{bn}(t_b|\theta_b) S_n(t_n|\theta_n) dG(\Theta) \tag{1}$$

where $G(\Theta)$ is the joint cumulative distribution function for $\Theta = \{\theta_{el}, \theta_n, \theta_b\}$. Following this approach we can write the likelihood function of the observed data.

3.3.4 The College Choice

So far we have assumed that college education was exogenous. A well-established literature however suggests that this is not the case (for example, Becker 1964). In our case college education can be related not only to "ability" in general but more specifically to "business talent." The framework described above can be easily modified to account for such endogeneity. We augment each individual's contribution to the likelihood function with a selection equation into college and rely on instrumental variables to achieve identification.

The selection equation simply describes the probability of getting a college degree at time t_{sc}. Such probability depends on a series of observables ($\hat{X}_{i,t}$), an unobserved term ($\theta_{i,sc}$) and a set of instruments (Z_i). Its formulation is similar to the rest of the model:

$$\lambda_{i,s}(t_{sc}|\theta_{i,sc}) = (1 + \exp(-y_{i,s}(t)))^{-1}$$

with

$$y_{i,s}(t) = \beta_s \hat{X}_{i,t} + \gamma_s Z_i + \theta_{i,s}$$

The set of controls contains variables also used in the description of transitions across working and non-working spells, such as race, age, intelligence, locus of control, marital status and parental background. The instrumental variables, as discussed above, are local unemployment rate at the time of high school graduation and distance from college interacted with parental background.

The contribution to the likelihood function of someone who got a college degree at time t_{sc} and then had the employment history discussed above is:

$$\int \prod_{c=1}^{t_{sc}-1} \left[1 - \lambda_{i,sc}(c_{sc}|\theta_{sc})\right] \lambda_{i,s}(t_{sc}|\theta_{sc}) f_{eb}(t_e|\theta_{el}) f_{bn}(t_b|\theta_b) S_n(t_n|\theta_n) d\hat{G}(\hat{\Theta}) \quad (2)$$

where $\hat{G}(\hat{\Theta})$ is the joint cumulative distribution function for $\hat{\Theta} = \{\theta_{el}, \theta_n, \theta_b, \theta_{sc}\}$.

4 Data

We rely on the National Longitudinal Survey of Youth - 1979 (NLSY79). The full sample includes 12,686 individuals designed to reflect a representative sample of the U.S. population. The NLSY79 contains very detailed data about

employment. We identify someone as being a business owner if they report being self-employed for most of a given year.[2] Education is measured by a dummy variable equal to one if the individual has a 4-years college degree and zero otherwise. Cognitive skills are measured by a test similar to the Armed Forces Qualification Test (AFQT), which is widely used in labor market studies. A measure of non-cognitive skills which we account for is "locus of control." This trait has been found to influence the choice to become self-employed (for example, Berlew, 1975). Locus of control measures the extent to which someone believes to be in control of the events that affect his or her life (Rotter, 1966).

Moving from self-employment to employment may not reflect failure but career advancement or change in preferences, so we cannot use this transition to study business success. However approximately one third individuals who leave self-employment move into non-employment. We define transitions from self-employment into non-employment as business failure. This is especially clear for those who remain non-employed for at least one year.

Male and female patterns of self-employment differs in terms of industry and the role of education (Macpherson, 1988; Simpson and Sproule, 1998).

[2]The Appendix contains a detailed description of how we constructed the employment status.

For this reason we analyse men and women separately. Since there are roughly twice as many male self-employed (Table 1), the precision of the estimates is lower for women.

4.1 College Data

We construct distance from college using data from National Center for Education Statistics (NCES) list of American four-years colleges. There are 2,966 colleges in the United States; for each college we collect information on legal status (public or private), geographical location, number of graduate and undergraduate students and other characteristics. The distance between the geographically closest college at time of high school graduation is defined as distance to college.

5 Results

5.1 College Education and Business Survival

Table 2 reports the main findings of this paper for the self-employed. Among men, without controlling for ability and selection, a college degree significantly reduces the probability of leaving a business spell into non-employment. Once

we add cognitive skills and control for selection we find that college has no statistically significant effect on business survival. Cognitive skills however have a large and statistically significant impact on survival. A one standard deviation increase in cognitive skills lengthen the average business spell by about 13% among men.

For women the effect of college on failure is not significantly different from zero even without controls, though the point estimate is negative. Once we add controls for cognitive skills and take into account selection using the model, the effect size of college is reduced and is virtually zero.

5.2 College Education and "Employment" Survival

Table 3 reports analogous results for the transition of employees into non-employment, which can be interpreted as losing one's job. Whereas a college degree has no effect on business survival, it increases "employee survival" for both men and women. The relationship remains statistically significant after controlling for ability and selection, though, perphaps not surprisingly, the effect size is reduced for both men and women. Interestingly, the effect of college on the risk of losing employment appears to be somewhat larger for men than for women. Cognitive ability reduces the probability of moving from employment to non-employment both with and without controls for selection.

The effect size of cognitive ability of remaining employed is large and similar in magnitude for both men and women.

The comparison of results in tables 2 and 3 suggests that a college education helps employees avoid non-employment, but does not help the self-employed. One interpretation of this finding is that college may teach skills such as the ability to work in teams or within a predetermined hierarchy that are more valuable for corporate employees than the self-employed.

5.3 Testing The Validity Of The Instruments

Table 4 reports the coefficients for the instrumental variables in the selection equation. The instruments seems to work as expected for males, even though the coefficient for proximity to college is estimated with a low degree of precision. The coefficient for proximity is also not significant for women and has the wrong sign. As a robustness check we have used a specification where only the deviation from unemployment is used as an instrument. The results were unchanged.

For our instruments to be valid it is necessary that they influence self-employment dynamics only through their impact on education. Possible concerns are that areas near colleges display higher rates of self-employment, or

families living close to colleges have higher personal wealth. Similarly, areas with higher relative unemployment at the time of high school graduation have higher self-employment rates. This does not appear to be the case (Table 5).

6 Conclusions

Studies that do not control for ability and selection generally find that the self-employed with a college degree have lower risk of business failure. We attempt to account for the endogeneity of both education and career choices by relying a competing risk model combined with instruments and controls for cognitive and non-cognitive abilities. Once selection and ability are controlled for, college no longer reduces the risk of moving from self-employment to non-employment. This indicates that previous studies may have overestimated the importance of college education for business survival by not sufficiently accounting for unobserved heterogeneity and endogeneity. In particular it seems as though college education captures the effect of cognitive skills - as measured by the AFQT - which have a positive and significant effect on the probability of avoiding non-employment.

A growing literature argues that the self-employed should be treated as conceptually distinct from "Schumpeterian" entrepreneurs (for example, Hurst

and Pugsley, 2010; Sanandaji, 2010). Self-employed individuals such as plumbers, "mom-and-pop" shop owners and dentists generally do not engage in innovation and rarely grow their firms. Both Schumpeterian entrepreneurs and self-employed individuals are important for the economy, though in different roles. The empirical model considered in this study is general and can be applied both to innovative and non-innovative business owners. Since our estimation applies the model to data on the self-employed our empirical results should be interpreted as the effect of college on the self-employed rather than entrepreneurs.

Interestingly a college degree significantly reduces the probability of employees transition into non-employment even after controlling for ability and selection. The fact that college helps employees but not the self-employed suggests that education affects the two groups in different ways. More research is required to understand what specific knowledge or skills are acquired during the college years that help people thrive in the corporate sector but do not affect their firms' success. We suspect the reasons might be found among the social, interpersonal and team-work skills acquired during the years on campus but leave to future articles to confirm our intuition or provide alternative explanations.

7 References

Arcidiacono, P., P. Bayer, and A. Hizmo. 2010. Beyond Signaling and Human Capital: Education and the Revelation of Ability. American Economic Journal: Applied Economics: 76-104.

Angrist, J. D., and A. B. Krueger. 1991. Does compulsory school attendance affect schooling and earnings?. The Quarterly Journal of Economics 106(4): 979-1014.

Becker, G. S. 1964. Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. University of Chicago Press, Chicago, IL.

Berlew, D. 1975. The nature of entrepreneurs. In Proceedings of Project ISEED (International Symposium on Entrepreneurship and Enterprise Development). Columbus, Ohio.

Boden Jr, R. J., and A. R. Nucci. 2000. On the survival prospects of men's and women's new business ventures. Journal of Business Venturing 15 (4): 347-362.

Bruderl, Josef, Peter Preisendorfer, and Rolf Ziegler. 1992. Survival

chances of newly founded business organizations. American sociological review: 227-242.

Card, D. 1995. Using geographic variation in college proximity to estimate the return to schooling in: Aspects of labour market behaviour: essays in honour of John Vanderkamp. Edited by L. N. Christofides, E.. K. Grant and R. Swidinsky. University of Toronto Press, Toronto, Canada.

Card, D. 1999. The Causal Effect of Education on Earnings. Handbook of Labor Economics, Volume 3, Ch. 30. Edited by O. Ashenfelter and D. Card. Elsevier Science.

Davidsson, P. and B. Honig. 2003. The Role of Social and Human Capital among Nascent Entrepreneurs. Journal of Business Venturing 18: 301-31.

Davis, S. J., J. Haltiwanger and S. Schuh, 1996, 'Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts. Small Business Economics 8: 297-315

De Brabander, D., J. Hellemans and C. Boone. 1999. Selection Pressure Induces a Shift Towards More Internal Scores on Rotter's I-E Locus of Control Scale. Current Research in Social Psychology 4(1): 103-12.

De Wit, G. and F. A. A. M. van Winden. 1989. An Empirical Analysis of

Self-Employment in the Netherlands. Small Business Economics 1: 263-72.

Duttweiler, P. C. 1984. The Internal Control Index: a Newly Developed Measure of Locus of Control. Educational and Pschological Measurement 44(2): 209-21.

Eberwein, C., Ham J. C. and R. J. LaLonde. 1997. The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data. The Review of Economic Studies 64(4): 655-682.

Evans, D. S. and L. S. Leighton. 1989. Some Empirical Aspects of Entrepreneurship. American Economic Review 79(3): 519-35.

Evans, D. S. and L. S. Leighton. 1989. Small Business Formation by Unemployed and Employed Workers. Small business Economics 2: 319-30.

Ganotakis, P. 2012. Founders' human capital and the performance of UK new technology based firms. Small Business Economics 39(2): 495-515.

Gimeno, F. J., T. B. Folta, A. C. Cooper and C. Y. Woo. 1997. Survival of the Fittest? Entrepreneurial Human Capital and the Persistence of Underperforming Firms. Administrative Science Quarterly 42: 750-83

Ham, J. C. and R. J. LaLonde. 1996. The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training. Econometrica 64(1): 175-205

Hartog, J., M. Van Praag, and J. Van Der Sluis. 2010. If you are so smart, why aren't you an entrepreneur? Returns to cognitive and social ability: Entrepreneurs versus employees." Journal of Economics & Management Strategy 19(4): 947–989.

Heckman, J. J. and B. Singer. 1984. Econometric Duration Analysis. Journal of Econometrics 24: 63-132.

Lin, Z., G. Picot, and J. Compton. 2000. The entry and exit dynamics of self-employment in Canada. Small Business Economics 15: 105-125.

Macpherson, D. A. 1988. Self-employment and Married Women. Economic Letters 28: 281-4.

Millan, J. M., E. Congregado, and Concepcion R. 2012. Determinants of self-employment survival in Europe. Small Business Economics 38(2): 231-258.

Rauch, A., and S. A. Rijsdijk. 2013. The Effects of General and Specific Human Capital on Long-Term Growth and Failure of Newly Founded

Businesses. Entrepreneurship Theory and Practice 37(4): 923–941.

Rotter, J. 1966. Generalized expectancies for internal versus external control of reinforcements. Psychological Monographs 80, Whole No. 609

Simpson, W. and R. Sproule. 1998. Econometric Analysis of Canadian Self-Employment Using SLID. Income and Labour Dynamics Working Paper No. 98-16

Spence, M. (1974). Job market signaling. Quarterly Journal of Economics, 87, 355-374.

Unger, J. M., A. Rauch, M. Frese, and N. Rosenbusch. 2011. Human capital and entrepreneurial success: A meta-analytical review. Journal of Business Venturing 26(3): 341-358.

Van der Sluis, J., M. van Praag and W. Vijverberg. 2008. Education and Entrepreneurship Selection and Performance: a Review of the Empirical Literature. Journal of Economic Surveys 22(5): 792-841.

Van Praag, M. 2003. Business survival and success of young small business owners: an empirical analysis. Small Business Economics 21(1): 1-17.

Van Praag, M., A. Van Witteloostuijn, and J. Van der Sluis. 2009. Returns for entrepreneurs versus employees: The effect of education and personal

control on the relative performance of entrepreneurs vis-a-vis wage employees. IZA DP N. 4628

Van Praag, M., A. Van Witteloostuijn, and J. Van der Sluis. 2013. The higher returns to formal education for entrepreneurs versus employees. Small Business Economics 40(2): 375-396.

Appendix

A Tables

	All	***Male***	***Female***
Observations	*12,686*	*6,394*	*6,292*
College	*23.52%*	*23.39%*	*23.65%*
Minorities	*19.33%*	*19.49%*	*19.16%*
Married	*59.15%*	*57.02%*	*61.36%*
Ever Owned a Business	*23.17%*	*27.17%*	*19.03%*
Business Owner	*6.95%*	*9.39%*	*4.45%*

Table 1: Comparison between men and women. All percentages are calculated as averages of the period 1988-2004 to refer to an adult population. In 1988 the youngest individuals is 24 years of age and virtually all the sample is out of school and either employed, self-employed or non-employed

	I	***II***	***III***	***IV***	***V***
			Panel A: Males		
College	*-0.33*			*0.15*	*-0.11*
	*(0.04)***			*(0.38)*	*(0.61)*
Intelligence		*-3.78*		*-3.94*	*-1.94*
		*(0.00)****		*(0.00)****	*(0.00)****
Self-Confidence			*0.51*	*0.35*	*0.77*
			(0.16)	*(0.34)*	*(0.05)***
			Panel B: ***Females***		
College	*-0.14*			*0.09*	*-0.03*
	(0.45)			*(0.65)*	*(0.90)*
Intelligence		*-2.35*		*-2.49*	*-1.01*
		*(0.00)****		*(0.00)****	*(0.22)*
Self-Confidence			*0.10*	*-0.07*	*0.35*
			(0.83)	*(0.88)*	*(0.47)*
Model	*No selec.*	*No selec.*	*No selec.*	*No selec.*	*Selec.*

Table 2: Human Capital and Business Survival. P-value in parenthesis. This table reports the coefficients of the variables on the leftmost column from the competing risks occupational model. Each regression also controls for marital status, number of kids, race, deviations of local unemployment rate from the mean, age, age squared and industry. Columns I-IV report results without correcting for endogeneity while column V reports the coefficient obtained correcting for the endogeneity of college education and entrepreneurship. Column VI reports the result for unincorporated businesses when correcting for endogeneity while column VII contains similar results for incorporated firms.

	I	***II***	***III***	***IV***	***V***
			Panel A: Males		
College	*-0.90*			*-0.52*	*-0.55*
	*(0.00)****			*(0.00)****	*(0.00)****
Intelligence		*-4.04*		*-3.70*	*-2.74*
		*(0.00)****		*(0.00)****	*(0.00)****
Self-Confidence			*0.05*	*-0.05*	*0.16*
			(0.64)	*(0.65)*	*(0.17)*
			Panel B: ***Females***		
College	*-0.52*			*-0.27*	*-0.27*
	*(0.00)****			*(0.00)****	*(0.00)****
Intelligence		*-3.84*		*-3.62*	*-2.76*
		*(0.00)****		*(0.00)****	*(0.00)****
Self-Confidence			*-0.01*	*-0.17*	*-0.02*
			(0.93)	*(0.12)*	*(0.87)*
Model	*No selec.*	*No selec.*	*No selec.*	*No selec.*	*Selec.*

Table 3: Human Capital and Business Creation. P-value in parenthesis. This table reports the coefficients of the variables on the leftmost column from the competing risks occupational model . Each regression also controls for marital status, number of kids, race, deviations of local unemployment rate from the mean, age, age squared and industry. Columns I-IV report results without correcting for endogeneity while column V reports the coefficient obtained correcting for the endogeneity of college education and entrepreneurship. Column VI reports the result for unincorporated businesses when correcting for endogeneity while column VII contains similar results for incorporated firms.

	Males	***Females***
Distance	*0.09*	*-0.09*
	(0.22)	*(0.15)*
Unemployment	*1.78*	*0.47*
	*(0.00)****	*(0.04)***

Table 4: Instrumental Variables in the college selection equation. P-value in parenthesis. This table reports the coefficients of the instrumental variables in the college selection equation. The equation also includes marital status, number of kids, race, age, age squared, a dummy for parental college education, a polynomial for duration, intelligence and self-confidence.

Panel A: Proximity to college				
	Males		*Females*	
	Wealth	***S.E Rate***	***Wealth***	***S.E Rate***
Close	*$58,559*	*5.6%*	*$57,260*	*2.5%*
	(954)	*(0.1%)*	*(897)*	*(0.01%)*
Far	*$88,233*	*6.6%*	*$88,531*	*2.9%*
	(1,742)	*(0.1%)*	*(1,747)*	*(0.01%)*
P-value	*0.00****	*0.00****	*0.00****	*0.00****
Panel B: Local Unemployment				
	Males		*Females*	
		S.E Rate		***S.E Rate***
High Unempl.		*5.2%*		*2.2%*
		(0.1%)		*(0.01%)*
Low Unempl.		*6.2%*		*2.9%*
		(0.1%)		*(0.01%)*
P-value		*0.00****		*0.00****

Table 5: Human Capital and Business Creation. P-value in parenthesis. This table reports the coefficients of the variables on the leftmost column from the competing risks occupational model. Each regression also controls for marital status, number of kids, race, deviations of local unemployment rate from the mean, age, age squared and industry. Columns I-IV report results without correcting for endogeneity while column V reports the coefficient obtained correcting for the endogeneity of college education and entrepreneurship. Column VI reports the result for unincorporated businesses when correcting for endogeneity while column VII contains similar results for incorporated firms.

B Construction of Yearly Employment Status

The discrete observation period is assumed to be a calendar year. Any construction of yearly employment and schooling status starting from weekly or monthly self-reported situation is somewhat arbitrary since an individual can be in several alternatives in a given year. There is no unequivocal solution to this problem. I followed the classification method proposed by Keane and Wolpin (1997) who used the same dataset to estimate a life-cycle model.

Every young man is assigned to one of four mutually exclusive states (employment, self-employment, non-employment or school) in the following hierarchical way. First I establish whether someone can be classified as employed, non-employed, self-employed or his/her status is missing for the year. a) **Missing Values, Employed or Non-employed**. If the weekly working status is missing for more than 2/3 of the weeks in one year, then the yearly status is missing. When weekly status is available for more than two thirds of the weeks then an individual is considered working if he/she reports doing so for more than two thirds of the non-missing weeks and averages at least 20 hours of work per week. Otherwise the yearly status is coded as "non-employment"[3]. b) **Self-Employed**. If an individual reports working as self-employed for more than half of the working weeks then he/she is considered self-employed for the

[3]Keane and Wolpin (1997) do a similar exercise but construct their employment variables looking at only nine weeks during the year. They do so for computational reasons. I do not have the same limitations so my working status uses all the information/weeks available. Keane and Wolpin also do not consider summer quarters to avoid picking up students' summer jobs. I calculate the working status with and without summer weeks. The correlation across individuals between the two definitions ranges between .9 in 1979 and .97 in 2003.

year.

Second, I establish whether someone classified as "Non-employed" is, in fact, in **school**. An individual is classified in school during the current calendar year if he/she is not already classified as employed or self-employed and one of the two following statements is true: a) he/she reports one more year of education the following calendar year *and* reports attending school at least during one month in the current calendar year; or b) he/she reports attending school for at least four months during current calendar year. The second part of this definition is meant to capture those individuals who spent most of their time in school but for whatever reason did not complete the grade[4].

[4]I decided to give priority to the employment information rather than the schooling attendance variable because the former seems to be more accurate. First, it is collected on a weekly basis rather than a monthly basis. Second, in order to be employed someone needs to work for more than 20 hours a week. Third, according to the rules of the NLSY79, it is enough to have attended school for just one day in order to be classified as in school for the entire month.