IESE Business School
University of Navarra

Working Paper

WP No 575

October, 2004

AN EMPIRICAL STUDY OF THE TRANSITION FROM PAID WORK TO SELF-EMPLOYMENT

S. Ramakrishna Velamuri*
Sankaran Venkataraman**

* Professor of Entrepreneurship, IESE

** Professor, Darden Graduate School of Business Administration

IESE Business School - Universidad de Navarra
Avda. Pearson, 21 - 08034 Barcelona. Tel.: (+34) 93 253 42 00 Fax: (+34) 93 253 43 43
Camino del Cerro del Águila, 3 (Ctra. de Castilla, km. 5,180) - 28023 Madrid. Tel.: (+34) 91 357 08 09 Fax: (+34) 91 357 29 13

AN EMPIRICAL STUDY OF THE TRANSITION FROM PAID WORK TO SELF-EMPLOYMENT

Abstract:

We explore the relationship between the probability of a transition from paid work to self-employment and three explanatory variables: paid income, predicted income, and income for ability. We use panel data for heads of households from the PSID SRC sample for eight pairs of years. Our results show that the relationship between paid income and self-employment is not linear. We then break up paid income into two components: a) predicted income based on human capital, demographic, and locational variables, and b) income for ability. Again, we find no linear relationship between self-employment and either predicted income or income for ability. We then test for curvilinear relationships between these three variables (i.e., paid income, predicted income, and income for ability) and the transition to self-employment. We find that individuals with low incomes are more likely to take up self-employment. Further, income for ability is a stronger predictor of the transition to self-employment than predicted income. We show that the relationship between ability and self-employment is U shaped: very low ability and very high ability individuals are more likely to take up self-employment than medium ability individuals. We use prospect theory to explain this result.

Keywords: Entrepreneurship, self-employment, opportunity costs, value creation

AN EMPIRICAL STUDY OF THE TRANSITION FROM PAID WORK TO SELF-EMPLOYMENT

Introduction

The decision by individuals to take up self-employment is one of the first steps that leads to the creation of many new firms, products, and services, and it is one of the two routes for the entrepreneurial process to proceed towards outcomes such as innovation and societal value creation, the other being corporate venturing. Carroll and Mosakowski (1987) argue that "while self-employment does not always generate new organizations, it is one important mechanism by which new organizations arise. Thus, organization theory may profit indirectly from future research in this area" (1987: p. 587). According to Shaver and Scott (1991), several pre-firm events contribute to the creation of new organizations. They explain the relevance of the different "events" as follows:

> "We agree that the actual founding of an organization is a discontinuous change, but do not believe that it is the only "event" to be explained. Nor is it likely that the founders of new business organizations would agree that they have done only *one* thing. Rather, founders would probably agree that they have done *many* things in order to get to the final one. Thus we believe that the discontinuous creation of a new venture is preceded by a series of prior discontinuous events, each of which can be considered a separate 'outcome'" (1991: 34).

We understand the transition from paid employment to self-employment to be one such discontinuous event, and study how it is influenced by income in paid work. Income in paid work has been used as a proxy for the opportunity cost of self-employment. Amit, Glosten and Muller (1993) pinpoint opportunity cost as an important gap in our understanding of self-employment.

> "We do not have models that examine the influence of opportunity costs on the willingness of would-be entrepreneurs to embark on highly uncertain projects. *Ceteris paribus*, it can be conjectured that equally able individuals with low opportunity costs are more likely to accept entrepreneurial projects. (This may perhaps, explain why relatively few MBAs are attracted to starting their careers as entrepreneurs. It means giving up a high salary)" (1993: 818).

Note: We gratefully acknowledge the financial support of the Batten Institute of the Darden School of Business Administration, University of Virginia, for carrying out this research. We wish to thank Paul Allison, Terrence Brown, Irene Duhaime, Alice De Koning, John Pepper, Scott Shane, and Malathi Velamuri for helpful comments on earlier drafts of this paper, and Prasanna Velamuri for assistance with data preparation. We also want to thank Tecla Loup of the Panel Study of Income Dynamics for several clarifications on the dataset, and Prasanna Velamuri for assistance with downloading data.

Amit, Muller and Cockburn (1995), using data from the Canadian Labor Market Survey (LMAS), found support for the hypothesis that the lower a person's opportunity cost (defined as wage earnings in the period immediately prior to the self-employment decision), the more likely she is to take up self-employment.

Several other studies have also found correlations between the probability of a transition from paid-employment to self-employment on the one hand, and income levels (Evans and Leighton, 1989), human capital (Rees and Shah, 1986), race (Bates, 1997; Borjas and Bronars, 1989), wealth (Bates, 1997; Evans and Jovanovic, 1989), and gender (Devine, 1994) on the other. We examine the theoretical arguments and empirical evidence for paid income as a predictor of the transition to self-employment.

A more nuanced understanding of who becomes an entrepreneur and under what circumstances can illuminate several issues of importance. First, there is the finding by Blau (1987) that the more developed the economy, the lower the percentage of individuals in non-agricultural self-employment. This finding holds both for longitudinal US data over a hundred years until 1970, as well as for a cross-sectional comparison of countries at varying stages of economic development. This suggests that the quality of entrepreneurship matters more than sheer numbers. Specifically, it would appear that it is in the interest of society for high human capital individuals to take up self-employment, and to bid low human capital individuals away from subsistence self-employment. Kangasharju and Pekkala (2001) have found that firms founded by highly educated individuals have higher growth potential. Second, the reversal of the US trend since the early 1970s of declining non-agricultural self-employment suggests that structural changes in the economies of the developed nations might favor industries in which economies of scale are not as important as in others (Blau, 1987). Another possible explanation for this reversal is the increase in the percentage of women and minorities in the labor force; the self-employment rate for both these groups has been rising steadily (Leung and Robinson, 2001). Third, since there is no market for self-employment (in the way in which there is a market for managers and other workers), individuals who wish to become self-employed must necessarily hire themselves (Knight, 1921). This leads to the question of why individuals in paid employment make the transition to self-employment, which is all the more interesting because very few individuals join self-employment when they join the labor force for the first time; the percentage of individuals becoming self-employed increases as a function of age (Rees and Shah, 1986).

We propose to study the self-employment decision, and more particularly, how it is influenced by income in paid work and the innate abilities of individuals. Our study examines the likelihood of self-employment using three measures. First, we use income in paid work to study if low-income individuals are more likely to take up self-employment. Through this first measure, we attempt to replicate the findings of Amit, Muller, and Cockburn (1995) and Evans and Leighton (1989). Second, we arrive at a predicted income measure using human capital, demographic, and locational variables. Third, we define income for ability as the difference between the actual income and the predicted income. Thus, income for ability is a residual measured in dollars that remains after human capital, demographic, and locational variables have been controlled for. Income for ability could be either positive or negative. A positive measure indicates a premium for ability; a negative measure indicates a discount for the lack of it. Our second and third measures simply split the paid income measure into predicted income and income for ability. We then examine which of the three measures – paid income, predicted income, or income for ability – is the strongest predictor of the transition from paid work to self-employment.

Section II briefly reviews the literature on the factors influencing the self-employment decision and formulates the hypotheses. Section III gives an overview of the methods and data source, defines the variables chosen, and outlines the regression models. Section IV presents the results, and Section V summarizes the conclusions.

Literature review and theoretical arguments

Many scholars have written on the likelihood of an individual's transition to self-employment using as explanatory variables income in paid work, human capital variables, age, race, and sex. Amit, Muller and Cockburn (1995), using data from the Canadian Labor Market Survey (LMAS), found support for the hypothesis that the lower a person's opportunity cost (defined as wage earnings in the period immediately prior to the self-employment decision), the more likely she is to take up self-employment. Consistent with the finding of Amit *et al.* (1995), Evans and Leighton (1989) reported from their study based on the National Longitudinal Study of Young Men (NLS) and the Current Population Surveys that lower paid wage workers are more likely to enter self-employment. Borjas and Bronars (1989) found that the self-employed are more likely to be college educated, a finding consistent with that of Rees and Shah (1986), which was based on data from the UK. Knudsen and McTavish (1989) also concluded that more educated individuals are more likely to take up self-employment.

We attempt to delve deeper into the positive relationship observed between self-employment and education, and the negative relationship observed between self-employment and income. In particular, we argue that theoretical arguments can be made in favor of a U-shaped relationship between income and self-employment, supporting both a negative relationship for low income individuals and a positive relationship for high income individuals.

Negative relationship between paid income and the self-employment decision

To support a negative relationship, one could say that the decision to take up self-employment on a full-time basis is a decision of considerable magnitude for those who are in paid employment. It requires sacrificing current income based on expectations of future income in the new career. A rational actor model would postulate that a paid worker analyzes all possible career alternatives, the payoffs of each alternative over time, and chooses the alternative whose net present value, weighted by the probability of its occurrence, is the highest. However, a complete list of career states, their probabilities, and their corresponding payoffs is not readily available to any paid worker. Even if this information were available, the costs for the worker of obtaining all the information and her limited computational ability would lead to her behaving in a boundedly rational way (Simon, 1955). Current income in paid employment is immediately available information that is a reasonable surrogate variable for the opportunity cost of becoming self-employed (see Amit, Muller, Cockburn, 1995: 97). Income in paid work is therefore the reservation wage against which the option of becoming self-employed is evaluated. Inertial forces favor the status quo, and paid employment is no exception. Prospect theory (Kahneman and Tversky, 1979) also provides support for this position by proposing that individuals are risk averse over gains, which means that an individual in paid work is more likely to overweight the prospect of sacrificing current income relative to foregoing a higher future income in self-employment. This phenomenon has also been referred to as the "endowment effect" (Knetsch and Sinden, 1984).

Positive relationship between paid income and the self-employment decision

On the other hand, several equally strong arguments can be put forward in favor of a positive relationship between paid income and the likelihood of self-employment. First, it can be argued that higher income individuals will see more valuable opportunities than lower income individuals, based on the finding that prior knowledge influences the quality of opportunities recognized. Shane (2000) found that prior knowledge is an important factor in determining the opportunities that different individuals will recognize and pursue. High-income individuals, who possess higher value knowledge, may perceive the expected value of their self-employment opportunities to be higher than the expected value of continuing in paid employment. Second, higher income individuals are more likely to have accumulated higher levels of wealth during their careers. This higher wealth could provide protection against possible downside losses in self-employment. Evans and Jovanovic (1989) found that capital market imperfections mean that the wealthy are more likely to take up self-employment. Bates (1997) found that one reason for the lower rates of self-employment entry for blacks relative to whites is their lower wealth holdings. The relationship between wealth and the likelihood of self-employment may well be curvilinear; at very high levels of wealth, the individual may not participate in the labor force at all. This is especially true if the individual has received the wealth in the form of an inheritance (see Holtz-Eakin, Joulfaian, and Rosen, 1993). Third, Blau (1987) suggested that higher marginal tax rates could be a key determinant for the self-employment decisions of high income individuals, because self-employment makes it easier to under-report income relative to paid work.

The above discussion paves the way for the following three hypotheses; the first two are competing hypotheses, whereas the third is an attempt to reconcile the two.

- *Hypothesis 1:* The probability of a transition from paid work to self-employment is negatively related to income in paid work.

- *Hypothesis 2:* The probability of a transition from paid work to self-employment is positively related to income in paid work.

- *Hypothesis 3:* Individuals with low and high incomes in paid work are more likely than individuals with middle incomes to make the transition from paid work to self-employment.

As already stated, the above discussion does nothing to disentangle the income effect on self-employment, which has been found to be negative (Amit, Cockburn, and Muller, 1995; Evans and Leighton, 1989), from the education effect, which has been found to be positive (Borjas and Bronars, 1989; Rees and Shah, 1986).

We find it puzzling that the relationship between paid income and the probability of self-employment has been found to be negative rather than U-shaped. There are two possible explanations for this. The first is that there are too few data points at high incomes, and this is biasing the parametric estimates to reflect the transitions at low and medium incomes. The other is that it is not income per se that matters, but the degree of dissatisfaction with it. Individuals at low and high incomes levels could be dissatisfied with their incomes for different reasons. We take up this discussion next.

The role of ability in the self-employment decision

Observation of the world of employment leads to an unequivocal conclusion: there are large differences in the income levels of individuals with the same level of human capital. This is so even for individuals in the same occupation and industry, in the same geographical location, and sometimes even in the same firm. We attribute these income differentials to differences in attributes such as the ability to get things done, energy levels, the ability to get along with people, motivation, commitment, etc. All these different characteristics, which are not captured by human capital variables, can be grouped under one general heading: ability. According to Behrman and Rosenzweig (1999), "ability has been used as the rubric for all unmeasured earnings endowments, which may include genetic endowments of ability, pre-school human capital investment, or motivation."

The focus to date of the research on ability has been on how to obtain accurate parametric estimates of the returns to education. The difficulty lies in separating the actual returns to education from the returns to innate abilities that are independent of education. It is recognized that education may be an endogenous variable with ability determining the amount of education related investment an individual will make. There are those who suggest that education may serve a function beyond that of productivity enhancement, as is suggested by human capital theory (Becker, 1993; Mincer, 1962; Schultz, 1961): it may serve as a signaling mechanism of an individual's ability (Spence, 1974). Thus, the ability variable may permit us to explore a lower level explanation for the probability of a transition to self-employment.

Blackburn and Neumark (1995) outline two measures that have been used for ability: first, explicit measures such as IQ test scores and parents' education have been used as proxies for ability. Second, data from identical (monozygotic) twins have been used to control for the ability factor. The conclusion one can draw from a review of the literature is that the concept of ability is of great theoretical importance, but no fully satisfactory method has yet been found of measuring it.

We offer a different approach to the ones that have been used so far, one that benefits from the fact that income estimation is not our primary research interest in this paper. We hypothesize that an individual's income in paid work can be explained by four broad constructs: first, human capital, which is made up of education, employment tenure, and labor market experience; second, demographic variables, based on the well-documented income penalties suffered in the US by non-whites and females; third, location variables, made up of occupation, industry, and state of residence; and fourth, innate abilities. Our model can therefore be written as

$$I = \alpha + \beta_1 H + \beta_2 D + \beta_3 L + \beta_4 A + \varepsilon,$$

where I is income in paid employment, H is a vector of human capital variables, D is a vector of demographic variables, L is a vector of locational variables, A is the ability measure, and ε is an error term uncorrelated with I. We assume that the market is efficient in rewarding the worker's productivity. Therefore, if all the variables except Ability are controlled for, then the residual (Income-Income') should contain income for ability and the error term. We take this residual to be a rough measure of the market's valuation of the ability of an individual. This residual can take both positive and negative values, reflecting an income premium for ability and an income penalty for the lack of it.

We recognize that there are problems with this measure. First, in predicting income, we have actually excluded ability, and thus all the parameter estimates could be biased. We offer a remedy to this problem by including a model with parents' education as proxy variables for ability; these variables have been used in the past to control for ability (for example, Card, 1995). The reader will note that there is not much difference in the parameter estimates between the two models. Second, we are unable to separate the error term from income for ability. While these difficulties are real, our position is that we are merely proposing that the residual be considered a rough measure of the market's valuation of an individual's ability.

We hypothesize that individuals with very low ability are more likely to take up self-employment. These individuals may simply lack the discipline to work under someone else's authority, or in teams, and may consequently be poorly rewarded. They may compare their low positions in their organizations and their low remunerations with those of higher ability individuals with similar human capital, and feel frustrated at the difference. They would thus be tempted to strike out on their own. On the other hand, individuals with very high ability are those with high energy levels, who have the ability to get things done, who have strong interpersonal skills, and who are creative problem solvers. These individuals may feel that, in spite of their above average remuneration, they could do better on their own. This is because they have to share with their principals a substantial percentage of their marginal contributions to their organizations. Finally, individuals of average ability are likely to be compensated in line with their human capital – they are therefore less likely to search for self-employment opportunities. Even if they come across such an opportunity, they are likely to assign lower weights to the expected future earnings from it relative to their current incomes in paid work (Kahneman and Tversky, 1979).

- *Hypothesis 4:* Low ability and high ability individuals are more likely than medium ability individuals to make the transition from paid work to self-employment.

Data, variables, and methods

We use data from the PSID SRC sample, a large longitudinal dataset begun in 1968 at the Institute for Survey Research. We focus on heads of households because they are the family members about whom the greatest amount of information is available. We take eight pairs of years: 1985-86, 1986-87, 1987-88, 1988-89, 1989-90, 1990-91, 1991-92, and 1992-93. The first year of each pair is year 1, and the second is year 2. We take 1985 as our starting point because some of the variables we are interested in are not available for earlier years. When the data preparation for this research was concluded, 1993 was the last year for which full data were available.

Generalized Least Squares Regressions

We first run two Generalized Least Squares (GLS) regression models, with paid income as the dependent variable. In Model 1, we use education, tenure, labor market experience, race, sex, occupation, industry, and state of residence as the independent variables. We use seven dummy variables for the eight years. To ensure against the omitted variable bias, we add two more variables – father's education and mother's education – as proxy variables for ability, in line with past practice (for example, Card, 1995). We then run

Model 2, without the two variables for parents' education. We present both models for the reader to see that there is not much difference in the parameter estimates between the two. We take the residual (Income – Income') from Model 2 as a rough measure of income for ability, to be used in the logistic regression models (see Goldstein (1995) for the practice of using residuals in multi-stage regressions).

Logistic Regressions

We then run four logistic regression models for clustered data using Generalized Estimation Equations, with a binary dependent variable for year 2 employment status (self-employment=1, paid work=0); in year 1 of each pair, all subjects are exclusively in paid work. We use year 1 variables as predictors of the self-employment choice in year 2.

In Model 1 we use paid income as the explanatory variable. In model 2 we use the two components of paid income from the GLS regression: predicted income and the residual (which is our income for ability). In order to test for a curvilinear relationship between these variables and the probability of transiting to self-employment, we set up four dummy coded variables for each of the three variables (that is, twelve variables in all) to represent the quintiles of the paid income, predicted income, and income for ability variables. The middle quintile (41st to 60th percentile) is the default quintile in all three cases. In model 3 we use the paid income quintiles. In model 4 we use the predicted income and income for ability quintiles.

There was a problem of quasi-complete separation with some of the dummy variables for occupation and state of residence. These variables were dropped from the respective models.

Dependent variable (self-employment in year 2): This variable is coded as 0 if the individual continued in paid employment in year 2, and as 1 if the individual made the transition to self-employment. All individuals in year 1 were in full time paid-employment.

Paid income: This is the total labor income of the individual. We use nominal incomes for multi-year data because we have 7 dummy variables for the eight years, which will control for all year-specific effects, including inflation.

Income quintiles: We have four dummy coded variables for the five income quintiles. Quintile 3 (1-20th percentile) is taken as the reference variable, and we thus report parametric estimates for quintiles 1, 2, 4, and 5.

Predicted income: This is the predicted income from the GLS model, which estimates paid income as a function of human capital, demographic, and locational variables. To protect against the omitted variable bias, we also use parents' education as proxies for ability.

Predicted income quintiles: To test for a non-linear relationship, we divide the predicted income residual into 5 quintiles, and obtain four dummy coded variables. The middle quintile (quintile 3) serves as our reference. Parametric estimates are reported for quintiles 1, 2, 4, and 5.

Income for ability: As explained earlier, income for ability is the residual between the actual income and the predicted income that we get from the GLS regression.

Income for ability quintiles: To test for a non-linear relationship, we divide income for ability into 5 quintiles, and obtain four dummy coded variables. The middle quintile (quintile 3) serves as our reference. Parametric estimates are reported for quintiles 1, 2, 4, and 5.

Control variables: In addition to the above, we control for several other variables. *Father's education* and *mother's education* are two variables that are coded identically: from 0 to 8, depending upon the levels of educational attainment of the parent. *Unemployed hours* is the total number of unemployed hours reported for year 1. *Children* is the number of children. *Jobunion* is dummy coded with a value of 1 if the paid job in year 1 was under union contract and 0 otherwise. *Tenure* is employment tenure in months. *Labmktex* is total full time labor market experience in years since the age of 18. *Tenuresq* and *Labmktexsq* are quadratic expressions of tenure and labmktex respectively to control for positive but declining returns to these two attributes (Ben-Porath, 1967). *Marital status* is coded 1 if the subject was married or cohabiting with a partner, and 0 otherwise. *Education* is years of education, with the values 0-16 representing the actual years of schooling, and 17 representing a maximum value for all graduate education. Thus, the mean of the education variable is somewhat understated in the dataset. *Age* is given in years, and only observations in the range of 18-65 are considered. *Race* is coded 1 for whites and 0 otherwise. *Sex* is coded 1 for females and 0 for males. *Father's occupation* refers to the usual occupation of the subject's father when the subject was growing up, and is coded 1 for self-employed and 0 otherwise. We then include four sets of dummy variables: 11 *occupation* variables for a total of 12 occupations at the 3-digit level; 7 *year* variables for a total of 8 years; 11 *industry* variables for a total of 12 industries at the 3-digit level; and 51 *state* variables for the 50 states, District of Columbia, and other territories like Virgin Islands, Puerto Rico, etc.

Results and Discussion

Descriptive statistics are given in Table 1. The correlation matrix is given in Table 2; it reports bivariate correlations for descriptive purposes only – the correlations do not take into consideration the higher within-cluster correlations relative to the across-cluster correlations. The results of the GLS regression predicting income as a function of human capital, sex, race, occupation, industry, state of residence, and parents' education are given in Table 3.

Table 4 gives the parametric estimates for the four logistic regression models. Model 1 includes paid income quintiles and the other control variables. Paid income is not correlated to self-employment. The explanatory variables that are statistically significant predictors are whether the job is under union contract (negative), Tenure (negative), Education (positive), and Race (whites more likely to become self-employed). The number of unemployed hours is a marginally significant predictor.

Model 2 includes predicted income and income for ability in place of paid income. Again, neither of the two variables explains the self-employment choice. It is interesting to compare the education and race variables in models 1 and 2. While they are significant at the $p<0.05$ and $p<0.01$ level respectively in Model 1, they cease to be significant in Model 2. This is particularly interesting given the orthogonality of education to our income for ability measure.

Model 3 includes income quintiles, with the middle quintile (41st to 60th percentile) serving as the base case. The objective of introducing quintiles is to test for a curvilinear relationship between income and the self-employment decision. Individuals in income quintile 1 are far more likely, and those in quintile 2 marginally more likely, to take up self-employment relative to those in quintile 3. The estimates for quintiles 4 and 5 are not statistically significant. Education and race are once again statistically significant predictors at the $p<0.01$ level.

Model 4 includes the predicted income and income for ability quintiles. Predicted income quintiles are not related to the self-employment decision. As far as income for ability is concerned, quintiles 1 and 5 are statistically significant predictors (at the $p<0.01$ level) of the transition to self-employment. Interestingly, education is again no longer significant in predicting self-employment. Race, on the other hand, is.

We find it interesting that in the two models in which predicted income and income for ability are introduced (Models 2 and 4) in place of paid income, education ceases to be a predictor of the self-employment decision. This could be because education is endogenous, and controlling for ability makes the education effect disappear. The two variables that are consistently strong predictors of self-employment across all four models are whether the paid job is under union contract and employment tenure.

People of very low and very high ability are more likely to take up self-employment than those of moderate ability. We hypothesize that low ability individuals are frustrated at their low incomes relative to their human capital. They may also be envious of individuals of similar human capital earning substantially more. People of very high ability are also more likely to take up self-employment, in spite of earning substantially more than the mean income for individuals with similar human capital. One possible explanation for this is that these very high ability individuals, in spite of being highly rewarded, may feel that they unnecessarily hand over a substantial percentage of their marginal contributions to their principals.

Conclusions

Using panel data, we have shown that income for ability is a better predictor than paid income of the likelihood that a person will take up self-employment. We have used the residual derived from the difference between actual income and predicted income as a rough measure of ability. Individuals earning substantially below their expected incomes are considered low-ability individuals, whereas individuals earning substantially above their expected incomes are considered high ability individuals.

In contrast to the suggestion made by Rees and Shah (1986), we assume that success in self-employment does not require different skills from those required in paid work. Individuals who show high abilities in paid work believe that they can successfully transfer these abilities to self-employment. Although individuals with very high abilities receive considerable rewards in paid work, their compensation is by definition lower than their marginal product. Those individuals for whom the difference between their marginal product and their income is the greatest will be more likely to take up self-employment. Individuals with low abilities are penalized by the market, such that their earnings are considerably below what they should be, given their human capital. We use prospect theory to explain why low and very high ability individuals are more likely to take up self-employment.

Our results are consistent with theories of entrepreneurship based on Austrian economics. For example, Shane (2000) posits that different individuals identify different opportunities based on their prior knowledge and experience. It can be argued that high ability individuals are more likely to be exposed to higher value opportunities, and this makes their transition to self-employment more likely. Our explanation is complementary to that of Shane (2000) in that we explain why a high ability individual might be more predisposed to identifying entrepreneurial opportunities.

Our results are also consistent with non-monetary explanations, such as the desire to take on the challenge of creating an economic institution, the desire for independence; it could be argued that high ability individuals who earn high incomes are more likely to be motivated by objectives other than purely monetary ones.

Whether the income residual from the GLS regression can be considered an approximate measure of income for ability or not, our findings are significant in two ways: first, that very high and very low values of the residual are very strong predictors of the transition to self-employment; and second, when the income is replaced by predicted income and residual income, education ceases to be significant in predicting self-employment. These findings, at the very least, merit a closer examination by scholars in the future.

References

Amit, R., E. Muller, and I. Cockburn, "Opportunity Costs and Entrepreneurial Activity," *Journal of Business Venturing* 10:1 (1995), 95-106.

Bates, T., *Race, Self-Employment, and Upward Mobility*, (Washington, D.C.: The Woodrow Wilson Center Press, 1997).

Becker, G. S., "A Theory of Marriages, Part (1)," *Journal of Political Economy* 81:4 (July-August 1973), 813-845.

Becker, G. S., *Human Capital*, 3rd edition, (Chicago: University of Chicago Press, 1993).

Behrman, J. R., and M. R. Rosenzweig, "'Ability' Biases in Schooling Returns and Twins: A Test and New Estimates," *Economics of Education Review* 18:2 (1999), 159-168.

Ben-Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," *The Journal of Political Economy* 75:4 (1967), 352-365.

Blackburn, M., and D. Newmark, "Are OLS Estimates of the Return to Schooling Biased Downward? Another Look," *Review of Economics and Statistics* 77:2 (May 1995), 217-230.

Blau, David. M., "A Time-Series Analysis of Self-Employment in the United States," *The Journal of Political Economy* 95 (1987), 445-467.

Borjas, G. J., and S. G. Bronars, "Consumer Discrimination and Self-Employment," *The Journal of Political Economy* 97:3 (1989), 581-605.

Card, D., "Using geographic variation in college proximity to estimate the return to schooling," in Lonis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, eds. *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*, 1995 (University of Toronto, Canada): 201-222.

Devine, T. J., "Changes in Wage-and-Salary Returns to Skill and the Recent Rise in Female Self-Employment," *American Economic Review* 82:2 (May 1994), 108-113.

Evans, D. S., and B. Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy* 97 (1989), 808-827.

Evans, D. S., and L. S. Leighton, "Some Empirical Aspects of Entrepreneurship," *The American Economic Review* 79:3 (1989), 519-535.

Goldstein, H., *Multilevel Statistical Models*, (Second Edition, Edward Arnold, London, 1995).

Holtz-Eakin, D., D. Joulfaian, and H. S. Rosen, "The Carnegie Conjecture: Some Empirical Evidence," *The Quarterly Journal of Economics* Vol. CVIII (May 1993), 413-436.

Kahneman, D., and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47:2 (1979), 263-291.

Kangasharju, A., and S. Pekkala, "Self-Selected Self-Employed: The Role of Education in Success," PTT Working Paper #38, Helsinki (2000).

Knetsch, J. L. and J. A. Sinden, "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value," *Quarterly Journal of Economics* 99:3 (1984), 507-521.

Knight, F., *Risk, Uncertainty, and Profit*, (Chicago: University of Chicago Press, 1921).

Knudsen, K. R., and D. G. McTavish, "Modelling interest in entrepreneurship: implication for business and development," in R. H. Brockhaus, N. C. Churchill, J. A. Katz, B. A. Kirchhoff, K. H. Vesper, and W. E. Wetzel, eds., *Frontiers of Entrepreneurship Research* (1989), 67-86.

Leung, D., and C. Robinson, "Explaining the Recent Rise in Self-Employment: Lifecycle, Cohort, and Aggregate Economy Effects," University of Western Ontario mimeo, (2001).

Mincer, J., "On the Job Training: Costs, Returns, and Some Implications," *Journal of Political Economy* 70:5 (1962), 50-79.

Rees, H., and A. Shah, "An Empirical Analysis of Self-Employment in the UK," *Journal of Applied Econometrics* 1 (1986), 95-108.

Schultz, T., "Investment in Human Capital," *American Economic Review* 51:1, (March 1961), 1-17.

Shane, S., "Prior knowledge and the discovery of entrepreneurial opportunities," *Organization Science* 11:4 (2000), 448-469.

Simon, H. A., "A Behavioral Model of rational Choice," *Quarterly Journal of Economics* 59, (1955), 99-118.

Spence, M., *Market Signalling*, (Cambridge: Harvard University Press, 1974).

Table 1. Descriptive Statistics

Number of valid observations	18,860			
Number of clusters	3,908			
Categorical Variables:				
Frequency of self-employed in year 2	524	(Missing Values 370)		
Job under union contract	3,955	(Missing Values 517)		
Marital Status (1=married or cohabiting)	13,061	(Missing Values 0)		
Race (1=white)	17,013	(Missing Values 43)		
Sex (1=Female)	3,405	(Missing Values 0)		
Father's occupation (1=self-employed)	673	(Missing Values 0)		
Continuous Variables:	Mean	Std. dev.	Min.	Max.
Paid income	30,428	23,251	0	550,000
Predicted income	29,837	12,281	-19,988	65,554
Income for ability	663.02*	19,347	-57,174	503,069
Unemployed Hours	45.21	187.10	0	2,080
Number of Children	0.98	1.14	0	7
Tenure (months)	99.44	100.65	0	564
Labor mkt. experience (yrs)	13.85	10.33	0	47
Education (yrs)	13.39	2.35	3	17

* Note that the mean residual of a GLS regression can be non-zero.

Table 2. Correlation Matrix

	Self-emp	Paid inc	Pred inc	Inc ability	Fath edu	Moth edu	Unemp hrs	Children	Job union	Tenure	Labmkt ex	Educ	Race	Sex	Marit stat
Self-emp															
Paid income	-0.0199***														
Predicted income	-0.0138*	0.5558***													
Income for ability	-0.00998	0.8496***	0.0338***												
Father's education	0.0287***	0.1439***	0.1902***	0.0540***											
Mother's education	0.0292***	0.1281***	0.1897***	0.0371***	0.6166**										
Unemployed hours	0.0258***	-0.1751***	-0.1610***	-0.1064***	-0.0482**	-0.0447**									
Children	-0.0065	0.0606***	0.0415***	0.0465***	-0.0503***	-0.0261**	0.0228**								
Job under union contract	-0.0554***	0.0190**	0.0531***	-0.0084	-0.0966***	-0.0833*	-0.0150**	0.0430***							
Tenure	-0.0621***	0.2475***	0.5063***	-0.0228***	-0.1887***	-0.1859***	-0.1163***	-0.0539**	0.2281***						
Labor market experience	-0.0128*	0.1347***	0.2371***	0.0107	-0.3508***	-0.3082**	-0.0716***	-0.1133**	0.1242***	0.5424**					
Education	0.0202***	0.3821***	0.6966***	0.0150**	0.4140***	0.3792***	-0.1103***	-0.0254**	-0.1309***	-0.0218***	-0.1444***				
Race	0.0241***	0.1377***	0.2391***	0.0117	0.1431***	0.1562**	-0.0608***	-0.0333**	-0.0207***	-0.0014	0.0060	0.1202***			
Sex	-0.0081	-0.2071***	-0.3451***	-0.0258***	0.0423***	0.0065	0.0161**	-0.1812**	-0.0348***	-0.0931**	-0.1203**	-0.0219**	-0.1364***		
Marital Status	-0.0085	0.2063***	0.2932***	0.0579***	-0.1048***	-0.0543***	-0.0415**	0.3575**	0.0547***	0.1598**	0.2098**	-0.0009	0.1226***	-0.6994**	
Father's occupation	0.0165*	0.0820***	0.0724***	0.0506***	-0.0262***	0.0242***	-0.0174*	-0.0334***	-0.0243**	0.0404**	0.0461***	0.0699***	0.0228**	0.0160*	-0.0019

* p<.10, ** p<.05, ***p<.01

Note: The correlations provided in this table do not control for the difference between within-cluster and across-cluster correlations. They have been provided for descriptive purposes only.

Table 3

GLS Regression - GEE Parameter Estimates
Dependent Variable = Paid Income

Parameter	Model 1 Estimate	Model 2 Estimate
Intercept	-27857.5***	-25934.6***
Tenure	67.1997***	65.6643***
Tenure squared	-0.0431*	-0.0367
Labor Market Experience	775.1564***	689.5555***
Labor Market Experience squared	-18.7449***	-17.4498***
Education	3071.368***	3168.512***
Race (white =1)	4607.615***	4978.80***
Sex (female = 1)	-7309.67***	-7157.63***
Father's education	425.0745**	—
Mother's education	199.08642	—

In addition to the variables listed above, dummy variables were used to control for year, occupation, industry, and state of residence. The estimates of these parameters have not been reported.

* p<.10
** p<.05
***p<.01

Note: The GEE tool used (PROC GENMOD in SAS with REPEATED statement for subject id) does not report a model R^2. The simple correlation between the observed and the predicted values of the dependent variable is 0.5528 for Model 1 and 0.5558 for Model 2, and that between the dependent variable and the residual is 0.8499 for Model 1 and 0.8496 for Model 2. These correlations do not control for higher within-cluster correlations.

Table 4

Logistic Regression - GEE Parameter Estimates
Dependent variable = Self-employed in year 2 (=1); Paid work in year 2 (=0)

Parameter	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Intercept	-6.7868***	-6.3328***	-7.5739***	-6.1446***
Paid income	-0.0000			
Paid Income Quintile 1			0.8521***	
Paid Income Quintile 2			0.3222*	
Paid Income Quintile 4			-0.1048	
Paid Income Quintile 5			0.2838	
Predicted income		0.0001		
Predicted Income Quintile 1				0.0403
Predicted Income Quintile 2				0.0759
Predicted Income Quintile 4				0.0316
Predicted Income Quintile 5				0.4073
Income for ability		-0.0000		
Income for ability Quintile 1				0.9282***
Income for ability Quintile 2				0.1209
Income for ability Quintile 4				-0.0309
Income for ability Quintile 5				0.4968***
Hours unemployed	0.0004*	0.0004*	0.0002	0.0002
Children	0.0226	0.0230	0.0214	0.0350
Job under Union Contract (Yes=1)	-0.7305***	-0.7315***	-0.6159***	-0.7064***
Tenure	-0.0059***	-0.0070*	-0.0052***	-0.0070***
Tenure squared	0.0000**	0.0000*	0.0000	0.0000*
Labor Market Experience	0.0244	-0.0128	0.0318*	0.0081
Labor Market Experience squared	-0.0003	0.0000	-0.0005	0.0001
Marital Status (Married=1)	-0.1295	-0.1287	-0.1213	-0.1369
Education in years	0.0802**	0.0271	0.0879***	0.0073
Race (White=1)	0.5496***	0.4655	0.6012***	0.4134**
Sex (Female=1)	-0.2178	-0.0979	-0.2723	-0.0706
Father's occupation (Self-employed=1)	0.3797	0.3816	0.3665	0.3288

The parameter estimates for occupation, year, industry, and state of residence have not been reported.

* p<.10, ** p<.05, ***p<.01

Model Goodness of Fit Criteria				
Deviance	4142.96	4142.84	4105.77	4081.54
Pearson chi-square	17936.06	17929.94	18184.91	18142.35
Log Likelihood	-2071.48	-2071.42	-2052.88	-2040.77