TI 2009-088/4
Tinbergen Institute Discussion Paper

Education and Entrepreneurial Choice: An Instrumental Variables Analysis

Jörn H. Block[1,2,3,4]
Lennart Hoogerheide[1,4,6]
Roy Thurik[1,3,5,6,7]

[1] *Erasmus University Rotterdam, The Netherlands;*
2 Technische Universität München, Germany;
[3] *Centre for Advanced Small Business Economics; The Netherlands;*
[4] *Econometric Institute; The Netherlands;*
[5] *EIM Business and Policy Research, Zoetermeer, The Netherlands;*
[6] *Tinbergen Institute, The Netherlands;*
[7] *Max Planck Institute of Economics, Germany.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Education and entrepreneurial choice: an instrumental variables analysis

Joern H. Block [a], Lennart Hoogerheide [b], Roy Thurik [c]

[a] Centre for Advanced Small Business Economics, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands and Technische Universität München, München, Germany. block@ese.eur.nl.

[b] Econometric Institute and Tinbergen Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. lhoogerheide@ese.eur.nl

[c] Centre for Advanced Small Business Economics, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands and EIM Business and Policy Research, P.O. Box 7001, 2701 AA Zoetermeer, the Netherlands. thurik@ese.eur.nl.

Abstract: Education is argued to be an important driver of the decision to start a business. The measurement of its influence, however, is difficult since it is considered to be an endogenous variable. This study accounts for this endogeneity by using an instrumental variables approach and a data set of more than ten thousand individuals from 27 European countries and the US. The effect of education on the decision to become self-employed is found to be strongly positive, much higher than the estimated effect in case no instrumental variables are used. That is, the higher the respondent's level of education, the greater the likelihood that he/she starts a business. Implications for entrepreneurship research and practice are discussed.

Keywords: Occupational choice, entrepreneurial choice; education, endogeneity, instrumental variables

First Version: October 2009

Current version: November 2010

FileName: Education and OccChoice Block Hoogerheide Thurik_V28

Save date: 11/23/2010 8:42:00 PM

JEL codes: C35, I20, J24, L26

Acknowledgements: Comments by David Ardia, Erik Canton, Per Davidsson, Adam Lederer, Mirjam van Praag, Dean Shepherd, André van Stel, Sebastien Tiggelovend, Wim Vijverberg, and Dinand Webbink are gratefully acknowledged. We acknowledge the help of Isabel Grilo of DG Enterprise and Industry of the European Commission. This note has been written in cooperation with the research program SCALES, carried out by EIM and financed by the Dutch Ministry of Economic Affairs. It benefitted from a short stay of Roy Thurik at EM-Lyon.

Corresponding author: Joern Block, Rotterdam School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands, block@ese.eur.nl, (tel. +31104081004).

Introduction

The determinants of entrepreneurial choice are widely researched (Evans and Jovanovic, 1989; Le, 1999; Lévesque et al., 2002; Wagner, 2007; Grilo and Thurik, 2008; Parker, 2009; Roper and Scott, 2009; Sena et al., 2010). Policy-makers are particularly interested in the effect of education since it can be influenced by policy measures (European Commission, 2003; OECD, 2009). Establishing its effect, however, is difficult due to endogeneity (Van der Sluis et al., 2008). That is, education appears as a causal variable in an econometric model while it is in fact correlated with the errors in the model. In general, this correlation can be caused by measurement errors or omitted variables.[1] Other causes include reverse causality, autoregression with autocorrelated errors and non-random samples (Kennedy, 2008).

In these situations the use of instrumental variables regressions (IV regressions) is a solution to isolate the causality (Angrist and Krueger, 1991).[2] Using IV regressions, Parker and Van Praag (2006) find that education is indeed endogenous to *entrepreneurial performance* and that it makes a difference whether or not IV methods are used. So far, however, we do not know of any study using IV regressions to analyze the effect of education on *entrepreneurial choice*. This is surprising, since entrepreneurial choice is widely examined in the literature. It is even more surprising since correcting for endogeneity of education is known to make a difference in the related area of *entrepreneurial performance* (Van der Sluis et al., 2008) and in those of the returns to education in general (Griliches and Mason, 1972; Blackburn and Neumark, 1993; Ashenfelter et al., 1999) and occupational choice in general (Siow, 1984; Zarkin, 1985).

While human capital theory predicts positive returns to education for both wage earners and entrepreneurs, no such theory exists for the effect of education on the entrepreneurial choice. In an ad hoc fashion both Gimeno et al. (1997) and Le (1999) argue that education may lead to skills that are useful for both entrepreneurs and wage earners and that no a priori effect of education on the choice between entrepreneurship and employment may be expected. Contrary to that, Davidsson and Honig (2003) argue that education provides individuals with increases in their cognitive abilities and is therefore positively associated with entrepreneurial discovery. Empirical exercises about the effect of education on entrepreneurial choice are not conclusive as the surveys of Grilo and Thurik (2008) and Parker (2009) show. But this may be due to the absence of a correction for endogeneity.

The present note has two aims. First, to analyze the effect of education on the entrepreneurial choice using IV methods and a large international data set and, second, to show to what degree not correcting for endogeneity leads to a bias.[3]

Our data set comprises of more than ten thousand individuals from 27 European countries and the US, who are either self-employed or in a paid employment job. We obtain two main findings: *first*, the effect of education on the decision to become self-employed is found to be strongly positive. The higher the respondent's level of education, the greater the likelihood that he/she starts a business. *Second*, our results show that a standard Probit or Logit model strongly underestimates the effect of education on entrepreneurial choice and leads to biased results. We suggest that this is the reason why earlier studies have found weak or insignificant results (Van der Sluis et al., 2008; Parker, 2009).

---

1 In the related literature on the determinants of *entrepreneurial performance* this can be justified as follows. Measurement errors in the observed education variable may push the estimated return to schooling towards zero, since they lead to variation in the education variable that has no effect on income. Further, education may be correlated with explanatory variables that are omitted. A typical omitted variable is unobserved 'ability'. Individuals with higher ability typically obtain higher education levels, but also earn higher income given a certain education level. This may lead to an overestimated return to schooling.

2 Alternative solutions are discussed by Heckman and Navarro-Lozano (2004).

3 For an analysis of asymptotic biases due to different types of specification errors in the probit model we refer to Yatchew and Griliches (1985).

Data and Method

To analyze the effect of education on entrepreneurial choice, we use data from the 2007 Flash Eurobarometer Survey on Entrepreneurship. The dataset has been used in a number of published studies (Grilo and Irigoyen, 2006; Grilo and Thurik, 2008; Stam et al., 2010; Van der Zwan et al., 2010) and contains detailed information on the respondents' employment status. We restrict the sample to those respondents who are either self-employed or in a paid employment job (10,962 obs.). We exclude respondents with solely domestic activities (1,678 obs.) or searching for a job (632 obs.), students (1,443 obs.), retirees (5,242 obs.), and respondents who refused to give an answer or do not fall in any of these categories (717 obs.). We lose some further observations due to missing values. The final dataset contains 10,397 observations.

Our dependent variable is a dummy variable, which indicates whether the participant is self-employed or not. Education is measured as the number of years which the participant spent receiving education. We include a number of commonly used control variables in the regression model such as gender or job experience (Grilo and Thurik, 2008). We also control for country effects. Table 1 describes the construction of the variables; Table 2 shows correlations and descriptive statistics.

Table 1 and 2 about here

To analyze the effect of education on the decision to become self-employed, we estimate both a standard Probit model and an IV Probit model. The IV model is estimated to account for the above discussed endogeneity issue associated with the education measure (Angrist and Krueger, 1991). There are two main groups of candidate instruments for education: family background variables and natural experiment variables such as changes or differences in compulsory schooling laws (Angrist and Krueger, 1991; Webbink, 2005; Hoogerheide et al., 2007). In general, family background variables are common, although not undisputed, instruments (Blackburn and Neumark, 1993; Parker and Van Praag, 2006). Other authors have used regional and legal variations in education as instruments, which are also not immune to criticism (Deaton, 2009; Heckman and Urzua, 2009; *The Economist*, 2009). Our dataset does not include the latter types of instruments. Hence, we rely on the first category and use the social class of the parents as instruments (e.g., blue collar vs. white collar).

We first test the *validity* of the instruments with the Amemiya-Lee-Newey minimum chi-square statistic (Amemiya, 1978; Newey, 1987; Lee, 1992). The null hypothesis of valid instruments is not rejected (p=0.146). That is, the data do not provide evidence against the overidentification restriction that is incorporated in the IV Probit model with multiple instruments for one possibly endogenous regressor. Hence, the instruments seem not to have a direct effect on the dependent variable: their only effect on the dependent variable seems to go via its effect on the endogenous explanatory variable.

Secondly, the instruments should be *statistically relevant* in the sense that they are correlated with the endogenous explanatory variable. Preferably, the instruments have a strong effect on the endogenous explanatory variable. Otherwise one is faced with the case of weak instruments in which it may be difficult to draw meaningful conclusions (Bound et al., 1995). In the education-income literature, famous weak instruments are the quarter of birth dummies of Angrist and Krueger (1991).[4] To test the strength of our instruments, we regress the supposedly troublesome variable education on our social class instruments and the controls. The *F* statistic

[4] Hoogerheide et al. (2007) show that only in a few southern states of USA, these instruments have a strong effect on education. For other regions, wide confidence intervals or posterior intervals are found for the return to schooling.

for the social class instruments is 17.76. This clearly exceeds 10.00 which is a widely used cut-off value to decide about the strength of an instrument (Kennedy, 2008).[5]

A caveat applies to IV methods as well: even if one is able to find valid and statistically relevant instruments, one should still be careful with the interpretation of the IV estimate. The IV estimate informs us only on the effect for those observations for which the instruments have power. This may be a small subgroup of the total population. For example, Hoogerheide et al. (2007) show that an estimate of return to education in the USA of Angrist and Krueger (1991) is determined by only a few Southern states. Several alternative approaches to IV exist. For example, Card (1999) provides an overview of studies using sibling and twin data to estimate the return to education and argues that omitted ability is eliminated when computing within family estimators. However, our dataset does not include the required observations on relatives. Alternatively, if one observes the same cross-sectional units over time and the endogeneity arises from time-invariant sources, fixed effects estimation could also eliminate endogeneity due to omitted variables. However, our data does not fit the required panel data framework.

Results

Table 3 shows the regression results.

Table 3 about here

The results regarding the effect of education on entrepreneurial choice are clear-cut: both in the standard Probit model and in the IV model, a positive effect of education regarding the decision to start a business is found. The IV model, however, shows a much stronger effect (β=0.014 in the standard Probit model[6]; β=0.137 in the IV model)[7, 8]. This strong difference in the size of the effects is explained by the fact that education is endogenous to entrepreneurial choice: estimating a standard Probit model underestimates the 'true' effect. The Wald-test of exogeneity is highly significant. The negative bias in the standard Probit model is also in line with the underestimation of the OLS estimator for the effect of education on wage (Angrist and Krueger, 1991).

There are two possible reasons for the underestimation of the effect in the standard Probit model. *First*, there may be omitted variables that are associated positively with education level and that have a negative effect on the decision to become self-employed (Griliches and Mason, 1972; Blackburn and Neumark, 1993). For example, a high education level may lead to attractive job opportunities in paid employment (e.g., an attractive position in a large corporation), which then increases the opportunity costs for starting a venture (Amit et al., 1995). It may also be that higher education creates awareness for the risks associated with entrepreneurship (Oosterbeek et al., 2010; Von Graevenitz et al., 2010), which then has a negative influence on the decision to start a business. Another possibility would be that institutions of higher education create a negative image or status of entrepreneurs as a group being exploiters of other people's work, which then has a negative effect on entrepreneurial choice (Begley and Tan, 2001). Finally, it may be

---

5 A value of 10 means that the IV bias is less than 10 percent of the OLS bias.

6 Using a standard Logit model yields β=0.024 ($p<0.001$).

7 The respective marginal effects are β=0.003 in the standard Probit model and β=0.023 in the IV model. Hence, an additional year of education increases the probability of becoming self-employed by 0.3% in the standard Probit model and 2.3% in the IV model. For the calculation, all dummy variables are set at zero (the modal value) and all continuous variables are set at their sample mean.

8 The education coefficient of the IV model would be β=0.135*** when including the 632 respondents *searching for a job* into the paid employment group.

that individuals with a strong preference for self-employment do not select into university education but rather prefer to attend a trade school or professional schools, which requires less years than university education. *Second*, if years of education is a poor proxy for the level of education[9] then measurement error drives the estimate for education in the standard Probit model towards zero or insignificance (similar to the linear model for which the effect of measurement errors is discussed by Griliches, 1977, and Angrist and Krueger, 1991).

The almost zero correlation between education and entrepreneurship may seem surprising at first sight (Table 2), given our finding of a significant effect of education on entrepreneurship in the IV model (Table 3). However, the correlation merely tells the story of the data without use of IV: that is, if this correlation were to be interpreted and hence 'trusted' as a proper indicator for the causality between education and entrepreneurship, we would not need IV in the first place. The correlation suffers from all the problems that the standard Probit estimator suffers from in the case of an endogenous explanatory variable: the -0.01 incorporates the influence of measurement errors, omitted variables, etc..

The results regarding the control variables are as expected (Grilo and Thurik, 2008; Parker, 2009; Sena et al., 2010). For example, male respondents have a higher likelihood of falling into the self-employment category (IV model: $\beta=0.388$, $p<0.001$). The effect of labor market experience is positive in its linear term and negative in its squared term. Country effects are important. For instance, the probability to become self-employed is higher in the US than in most European countries. But also among European countries there is wide disparity: the highest probability exists in Greece ($\beta=0.825$, $p<0.001$); the lowest probability exists in Denmark ($\beta=-0.454$, $p<0.001$). An F-test on joint significance of the country variables shows a significant result. The detailed discussion and interpretation of country differences is beyond the scope of the present note and has been discussed in prior research (Grilo and Thurik, 2005; Grilo and Irigoyen, 2006).

Discussion and Conclusion

The advent of the knowledge economy together with the recognition that such an economy requires a prominent entrepreneurial sector (Audretsch and Thurik, 2001) produced many studies regarding the effect of education on entrepreneurial choice and performance (Van der Sluis et al., 2008). Moreover, of the many factors known to influence entrepreneurial choice and performance (Evans and Jovanovic, 1989; Le, 1999; Lévesque et al., 2002; Grilo and Thurik, 2008; Parker, 2009; Roper and Scott, 2009; Sena et al., 2010) education is popular among politicians since it can be influenced. Our note contributes to this literature by estimating an IV model to explain the causal effect of education on entrepreneurial choice. We show that education appears to be an endogenous variable regarding the decision to become self-employed, which is why an IV model is needed to estimate its effect. Using such a model, we then show that a higher level of education increases the likelihood of becoming self-employed. Our data set comprises of more than ten thousand individuals from 27 European countries and the US, who are either self-employed or in a paid employment job.

These two main results have a number of implications for both method and practice: *first*, the popularity among politicians to promote education as an important driver of economic growth is supported by the effect that education promotes entrepreneurship which itself is a driver of economic growth (Thurik et al., 2008). *Second*, our results show that a standard Probit (or Logit) model shall not be used to estimate the effect of education, since it tends to underestimate the effect of education. An IV approach might be a solution to find the 'true' effect. In that respect, entrepreneurial choice does not differ from other educational outcome variables such as wage (Angrist and Krueger, 1991; Van Praag and Van der Sluis, 2004; Webbink, 2005).

[9] For example, years of education as a measure does not account for the *quality* of education (Link, 1973).

The results of our note offer several interesting avenues for further research. A particularly promising avenue concerns the discussion of necessity, opportunity and involuntary entrepreneurship (Block and Koellinger, 2009; Block and Wagner, 2010; Kautonen et al., 2010). Entrepreneurs who are 'pushed' into entrepreneurship might fall into a low education subgroup, in which case the effect of education on entrepreneurial choice would be negative. Another avenue of further research would be to analyze whether a higher level of education increases the preference for self-employment as a means to obtain non-monetary benefits (e.g., more flexibility or independence). Another avenue would be to use a more comprehensive dataset which includes more information about the individual's labor market status. Such a dataset would allow estimating the effect of education in a multinomial model, in which not only self-employment and employment exist as alternatives but also non-employment and unemployment (Grilo and Thurik, 2008). Lastly, it would be interesting to analyze whether the positive effect of education on entrepreneurial choice holds for all modes of entry into entrepreneurship (e.g. new venture start versus business takeover) and all countries alike (e.g., developing versus industrialized countries).

References

Amemiya T (1978) The estimation of a simultaneous equation generalized probit model. *Econometrica* 46(5): 1193-1205.

Amit R, Mueller E and Cockburn I (1995) Opportunity costs and entrepreneurial activity. *Journal of Business Venturing* 10(2): 95-106.

Angrist JD and Krueger AB (1991) Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106(4): 979-1014.

Ashenfelter O, Harmon C and Oosterbeek H (1999) A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics* 6(4): 453-470.

Audretsch DB and Thurik AR (2001) What is new about the new economy: sources of growth in the managed and entrepreneurial economies. *Industrial and Corporate Change* 10(1): 267-315.

Begley TM and Tan WL (2001) The socio-cultural environment for entrepreneurship: a comparison between East Asian and Anglo-Saxon Countries. *Journal of International Business Studies* 32(3): 537-553.

Blackburn ML and Neumark D (1993) Omitted-ability bias and the increase in the return to schooling. *Journal of Labor Economics* 11(3): 521-544.

Block J and Koellinger P (2009) I can't get no satisfaction—Necessity entrepreneurship and procedural utility. *Kyklos* 62(2): 191-209.

Block J and Wagner M (2010) Necessity and opportunity entrepreneurs in Germany: characteristics and earnings differentials. *Schmalenbach Business Review* 62(2): 154-174.

Bound J, Jaeger DA and Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90(430): 443-450.

Card D (1999) The causal effect of education on earnings. In *Handbook of Labor Economics*, *Volume 3*, edited by O. Ashenfelter and D. Card: 1801-1863.

Davidsson P and Honig B (2003) The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing* 18(3): 301-331.

Deaton AS (2009) Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development. NBER working paper 14690.

European Commission (2003) *Green Paper: Entrepreneurship in Europe*. Brussels: Directorate General of Enterprise and Industry.

Evans, DS and Jovanovic B (1989) An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy* 97(4): 808-827.

Gimeno J, Folta TB, Cooper AC and Woo CY (1997) Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative Science Quarterly* 42(4): 750-783.
Griliches Z and Mason WM (1972) Education, income, and ability. *Journal of Political Economy* 80(3): S74-S103.
Griliches Z (1977) Estimating the returns to schooling: some econometric problems. *Econometrica* 45(1): 1-22.
Grilo I and Irigoyen JM (2006) Entrepreneurship in the EU: to wish and not to be. *Small Business Economics* 26(4): 305-318.
Grilo I and Thurik AR (2005) Latent and actual entrepreneurship in Europe and the US: some recent developments. *International Entrepreneurship and Management Journal* 1(4): 441-459.
Grilo I and Thurik AR (2008) Determinants of entrepreneurial engagement levels in Europe and the US. *Industrial and Corporate Change* 17(6): 1113-1145.
Heckman JJ and Navarro-Lozano S (2004) Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics* 86(1): 30-57.
Heckman JJ and Urzua S (2009) Comparing IV with structural models: what simple IV can and cannot identify. NBER working paper 14706.
Hoogerheide LF, Kleibergen F and Van Dijk HK (2007) Natural conjugate priors for the instrumental variables regression model applied to the Angrist–Krueger data. *Journal of Econometrics* 138(1): 63-103.
Kautonen T, Down S, Welter F, Vainio P, Palmroos J, Althoff, K. and Kolb S (2010) "Involuntary self-employment" as a public policy issue: a cross-country European review, *International Journal of Entrepreneurial Behaviour & Research* 16(2): 112-129.
Kennedy P (2008) *A Guide to Econometrics*. 6th edition. Blackwell Publishing: Oxford.
Le AT (1999) Empirical studies of self-employment. *Journal of Economic Surveys* 13(4): 381-416.
Lee LF (1992) Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables. *Econometric Reviews* 11(3): 319-328.
Lévesque M, Shepherd DA and Douglas EJ (2002) Employment or self-employment: a dynamic utility-maximizing model. *Journal of Business Venturing* 17(3): 189-210.
Link CR (1973) The quantity and quality of education and their influence on earnings: the case of chemical engineers. *Review of Economics and Statistics* 55(2): 241-247.
Newey WK (1987) Efficient estimation of limited dependent variable models with endogeneous explanatory variables. *Journal of Econometrics* 36(3): 231–250.
OECD (2009) *Education at a Glance*. Paris.
Oosterbeek H, Van Praag M and Ijsselstein A (2010) The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review* 54: 442-454.
Parker SC (2009) *The Economics of Entrepreneurship, second edition*. Cambridge University Press: Cambridge.
Parker SC and Van Praag CM (2006) Schooling, capital constraints and entrepreneurial performance: the endogenous triangle. *Journal of Business and Economic Statistics* 24(4): 416-431.
Roper S and Scott JM (2009) Perceived financial barriers and the start-up decision. An econometric analysis of gender differences using GEM data. *International Small Business Journal* 27(2): 149-171.
Sena V, Scott J and Roper S (2010) Gender, borrowing patterns and self-employment: some evidence for England. *Small Business Economics*, DOI 10.1007/s11187-010-9272-9.
Siow A. (1984) Occupational choice under uncertainty. *Econometrica* 52(3): 631-645.
Stam E, Thurik AR and Van der Zwan P (2010), Entrepreneurial exit in real and imagined markets. *Industrial and Corporate Change* 19(4): 1109-1139.
The Economist (printed edition) (2009) Sources: cause and defect. August 13.

Thurik AR, Carree MA, Van Stel A and Audretsch DB (2008) Does self-employment reduce unemployment? *Journal of Business Venturing* 23(6): 673-686.
Von Graevenitz G, Harhoff D and Weber R (2010) The effects of entrepreneurship education. *Journal of Economic Behavior & Organization* 76(1): 90-112.
Van der Zwan P, Thurik, AR and Grilo I (2010) The entrepreneurial ladder and its determinants. *Applied Economics* 42(17): 2183 - 2191.
Van Praag CM and Van der Sluis J (2004) Economic returns to education for entrepreneurs: the development of a neglected child in the family of economics of education? *Swedish Economic Policy Review* 11(2): 183-226.
Van der Sluis J, Van Praag CM and Vijverberg, W (2008) Education and entrepreneurship selection and performance: a review of the empirical literature. *Journal of Economic Surveys* 22(5): 795-841.
Wagner J. (2007) What a difference a Y makes - female and male nascent entrepreneurs in Germany. *Small Business Economics* 28(1): 1-21.
Webbink D. (2005) Causal effects in education. *Journal of Economic Surveys* 19(4): 535-560.
Yatchew A and Griliches Z (1985) Specification error in probit models. *Review of Economics and Statistics* 67(1): 134-139.
Zarkin GA (1985) Occupational choice: an application to the market for public school teachers. *Quarterly Journal of Economics* 100(2): 409-446.

Table 1: Description of variables

Variable	Description
Dependent variable	
Self-employment	Dummy = 1 if respondent is self-employed
Variable of interest	
Education	Number of years the respondent has been in full-time education
Instruments [1]	
Father was/is white collar	Dummy = 1 if father of respondent had/has a white collar job
Father was/is blue collar	Dummy = 1 if father of respondent had/has a blue collar job
Father was/is civil servant	Dummy = 1 if father of respondent was/is civil servant
Father was/is without professional activity	Dummy = 1 if father of respondent was/is without professional activity
Mother was/is white collar	Dummy = 1 if mother of respondent had/has a white collar job
Mother was/is blue collar	Dummy = 1 if mother of respondent had/has a blue collar job
Mother was/is civil servant	Dummy = 1 if mother of respondent was/is civil servant
Mother was/is without professional activity	Dummy = 1 if mother of respondent was/is without professional activity
Control variables	
Labor market experience	Age of the respondent minus age when stopped full time education
Male	Dummy = 1 if respondent is male
Father was/is self-employed	Dummy = 1 if father of respondent was/is self-employed
Mother was/is self-employed	Dummy = 1 if mother of respondent was/is self-employed
Rural region	Dummy = 1 if respondent lives in a rural region
Metropolitan region	Dummy = 1 if respondent lives in a metropolitan region
Country dummies	28 Country indicator variables (Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, UK, US)

Note: [1] The instruments do not sum up to 1, since the response categories 'father/mother was/is self-employed' and 'don't know/no answer' are not used as instruments.

Table 2: Descriptive statistics and correlations

		Min.	Max	Mean	Median	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Self-employment	0	1	0.20	0																
2	Education	1	75	13.66	12	-0.01															
3	Labor market experience	0	81	23.02	23	0.11	-0.42														
4	Male	0	1	0.44	0	0.14	0.01	0.02													
5	Father was/is self-employed	0	1	0.24	0	0.16	0.02	0.04	0.03												
6	Mother was/is self-employed	0	1	0.10	0	0.10	-0.00	0.02	-0.01	0.41											
7	Rural region	0	1	0.35	0	0.04	-0.12	0.05	-0.01	0.05	0.05										
8	Metropolitan region	0	1	0.23	0	-0.01	0.14	-0.05	0.01	-0.04	-0.03	-0.40									
9	Country is US	0	1	0.05	0	0.05	0.12	0.00	0.05	0.00	-0.01	-0.02	0.09								
	Instruments																				
10	Father was/is white collar	0	1	0.17	0	-0.02	0.07	-0.08	-0.01	-0.25	-0.10	-0.03	0.05	0.05							
11	Father was/is blue collar	0	1	0.33	0	-0.09	-0.10	0.05	-0.00	-0.39	-0.16	0.04	-0.05	-0.01	-0.32						
12	Father was/is civil servant	0	1	0.18	0	-0.03	0.05	-0.02	-0.01	-0.26	-0.11	-0.09	0.05	-0.05	-0.21	-0.33					
13	Father was/is without professional activity	0	1	0.04	0	-0.01	-0.04	-0.02	-0.01	-0.12	-0.05	0.02	-0.02	-0.01	-0.10	-0.15	-0.10				
14	Mother was/is white collar	0	1	0.14	0	-0.02	0.09	-0.17	-0.03	-0.07	-0.14	-0.02	0.06	0.06	0.31	-0.10	-0.08	-0.06			
15	Mother was/is blue collar	0	1	0.19	0	-0.07	-0.07	0.01	-0.03	-0.16	-0.16	0.01	-0.01	0.00	-0.10	0.36	-0.12	-0.06	-0.19		
16	Mother was/is civil servant	0	1	0.14	0	-0.04	0.08	-0.11	-0.03	-0.11	-0.14	-0.08	0.07	-0.05	-0.06	-0.13	0.38	-0.06	-0.16	-0.19	
17	Mother was/is without professional activity	0	1	0.41	0	0.03	-0.06	0.17	0.06	0.02	-0.28	0.03	-0.06	-0.00	-0.03	-0.00	-0.04	0.16	-0.33	-0.40	-0.33

Notes: N=10,397 obs.; All correlations above r=0.02 have a p-value less than 0.05; We checked estimation results omitting 'outliers' (e.g. observations with education over 30 years): changes in results are minor; To save space, we have only included the US and not the 27 European countries; In the multivariate analysis, all country dummies (except for the US as a reference group) are included.

Table 3: Results of standard Probit regression and instrumental variables Probit regression
Dependent variable: Individual is self-employed

	Standard Probit Regression		Instrumental Variables Probit Regression (two step) [1]	
Variables	**Coefficient (SE)**		**Coefficient (SE)**	
Education [a, b]	0.014 (0.003)	***	0.134 (0.029)	***
Labour market experience/10	0.137 (0.038)	***	0.475 (0.096)	***
(Labour market experience/10)^2	0.001 (0.010)		-0.023 (0.010)	*
Male	0.392 (0.030)	***	0.388 (0.029)	***
Father was/is self-employed	0.347 (0.037)	***	0.290 (0.035)	***
Mother was/is self-employed	0.178 (0.050)	***	0.203 (0.064)	***
Rural region [c]	0.169 (0.035)	***	0.279 (0.048)	***
Metropolitan region [c]	0.051 (0.040)		-0.042 (0.054)	
Country dummies [d]	27 categories (p<0.001)		27 categories (p<0.001)	
Intercept	-1.999 (0.151)	***	-4.183 (0.642)	***
N	10,397		10,397	
Minus Log pseudolikelihood	4656.61			
Pseudo R^2	0.083			
Wald Chi² (df)	765.11 (35) ***		683.44 (35) ***	

* p<0.05 ** p<0.01 *** p<0.001
SE=robust standard errors (standard Probit regression); bootstrapped standard errors (Instrumental variables Probit Regression)

Notes:
[a] Instruments for education: 'father was/is white collar', 'father was/is blue collar', 'father was/is civil servant', 'father was/is without professional activity', 'mother was/is white collar', 'mother was/is blue collar', 'mother was/is civil servant', 'mother was/is without professional activity' (F-test for significance of the instruments in the regression of education: F(8, 10,392)=17.76***). The regression of education includes the instruments and the control variables (see Table 1).
Wald-test of exogeneity: p<0.001
Validity of the instruments: Amemiya-Lee-Newey minimum chi² statistic: 10.837 (p=0.146)
[b] When excluding outliers (education is more than 30 years), the coefficients are β=0.131 *** (IV model) and β=0.014 *** (Standard Probit Model). We also tested for a non-linear effect of education on entrepreneurial choice but found no evidence of such an effect.
[c] Reference category is 'other town/urban centre'.
[d] Reference category is 'US'.