The Returns to Education in Entrepreneurship: Heterogeneity and Non-Linearities*

JENS IVERSEN,
NIKOLAJ MALCHOW-MØLLER, AND
ANDERS SØRENSEN

October 15, 2010

Abstract

The returns to education in self-employment are addressed in four different specifications of the relationship between log income and years of schooling. The specifications range from a standard Mincer equation with a constant percentage increase in income to an additional year of schooling to the most flexible specification with dummy variables for the different number of years of schooling split into different types of education. Based on the more flexible specifications, important non-linearities and heterogeneity in the returns to education in self-employment are found. These results are robust across different estimation methods: OLS; Heckit correction models to handle sample selection; and IV to deal with the potential endogeneity

*Acknowledgements: We thank participants at the CEBR conference on "Entrepreneurship: Occupational Choice and Financing" and at the Arne Ryde symposium "The Economics of Higher Education and the Education of Economists" for helpful comments. Jonas Helth Lønborg provided efficient research assistance. We also thank the National Agency for Enterprise and Housing for financial support for this project. Sørensen gratefully acknowledges financial support from Tuborgfondet. The usual disclaimer applies.
Iversen and Malchow-Møller: Department of Business and Economics, University of Southern Denmark. Sørensen: Department of Economics, Copenhagen Business School. All: Centre for Economic and Business Research, Copenhagen Business School.

of years of schooling. Moreover, the results are insensitive to the use of different sample years, different definitions of self-employment, and different income measures for the self-employed.

1 Introduction

According to the human capital theory pioneered by Becker (1963), education is an investment of current resources (time and money) in exchange for future pecuniary returns in the form of higher earnings.[1] Politically and academically, there has long been a huge interest in estimating the private (and social) pecuniary returns to education. Card (1999), Harmon *et al.* (2003) and Heckman *et al.* (2006) are all examples of recent reviews of the empirical literature estimating returns to education for wage employed.[2] The relationship between education and earnings has also received considerable interest in the entrepreneurship literature; see Parker (2004) and van der Sluis *et al.* (2008) for recent surveys.

The main work horse for estimating returns to education is the so-called Mincer equation – derived by Jacob Mincer in 1974 – which assumes a log-linear relationship between years of schooling and earnings. That is, an extra year of schooling yields a constant relative increase in future earnings of α %, where α is the coefficient to years of schooling.

In the literature on returns to education in wage work, non-linearities in the relationship between years of schooling and earnings have been documented by Heckman *et al.* (2006) and other studies cited therein. Furthermore, results from Chevalier *et al.* (2002) and Walker and Zhu (2001) indicate variations in returns to different types of higher education in wage work. Other studies, however, find that the relationship between years of schooling and log earnings is well described by a linear function as in the original Mincer formulation; see, *e.g.*, Harmon *et al.*, (2003) and Card (1999).

In the entrepreneurship and self-employment literature, non-linearities and heterogeneity across types of education have not received much attention. Most studies rely on the classical log-linear Mincer specification with years of schooling as the independent variable, while some studies include one or at most a few dummies for educational attainment; see van der Sluis *et al.*

[1] The importance of non-pecuniary returns such as a higher social status or the possibility of finding a more interesting job has also been emphasised; see, e.g, Becker (1964), Heckman (1976), and Lazear (1977). However, in the present paper we focus (as most of the studies in the literature) on the pecuniary returns.

[2] Strictly speaking, most of the studies estimate the growth rate of market earnings with years of schooling. Only under very specific conditions can this be interpreted as the internal rate of return to education. See Heckman et al. (2005) for a discussion of this. However, as it is common in the literature to refer to these estimates as "rates of return", we will maintain this terminology in the present paper.

(2008).

However, a recent short paper by Iversen *et al.* (2010) indicates that there is likely to be important non-linearities and heterogeneity in the returns to education for self-employed and entrepreneurs. Using simple OLS regressions, a highly non-linear relationship is detected between years of schooling and log earnings with very low returns in self-employment to most levels of education, and with considerable variation across different types of (higher) education as well. This indicates that non-linearities and heterogeneity in the returns are likely to be much more important for self-employed than for wage workers.

As this issue has been largely neglected in the entrepreneurship literature, the purpose of the present paper is to provide a deeper investigation of these preliminary findings. Are the results robust to the use of instrumental variables techniques to deal with the potential endogeneity of schooling, and are they influenced by sample selection? These are some of the questions posed in the present paper. We also investigate whether the findings are sensitive to the choice of sample year, the definition of the self-employed, the income measure used, and the industry of the self-employed.

For this purpose, we use register data containing detailed information on educational attainment, earnings and occupations for all Danish residents. Using these data and a standard Mincer specification, we find the same overall return as in previous studies; see, *e.g.*, van der Sluis *et al.* (2008). However, when we introduce more flexible specifications of the relationship between education and income, by allowing for different return rates across different levels, we confirm the findings from Iversen *et al.* (2010) as we strongly reject the log-linear relationship between years of schooling and income. Instead, using dummy variables for educational attainment, we find that only individuals with 18 or more years of schooling experience substantial returns to education. We also find substantial heterogeneity in the returns across subject areas for a given educational length.

As often stressed in the literature, OLS estimates may be inconsistent as (i) the measure of education is likely to be endogenous, typically due to the presence of unobserved individual ability affecting both education and earnings; and (ii) sample selection bias, as the sample of self-employed used is not a random draw from the population.

In order to deal with the first problem, family background variables have often been used in the literature as instruments for the educational attainment of an individual. In the present paper, we follow this approach and use the educational attainment of the parents and/or the spouse as instruments. While we believe that these are the best available instruments, we remain sceptical about their validity as we discuss below. However, using the instruments either confirms the non-linear relationship between years of schooling and earnings or leads to insignificant coefficient estimates.

A Heckman correction model is used to deal with the second problem. Using the amount of parental experience from self-employment as an extra regressor in the probit modelling the selection into self-employment, we find indications of a sample selection problem. Correcting for this, however, does not substantially affect the estimated effects of education.

Furthermore, as we shall discuss at length below, the empirical definitions of the self-employed and their income are far from trivial. We therefore use alternative definitions of both to check the sensitivity of our results to these. Although the alternative definitions change the sample and the dependent variable substantially, they only have minor effects on the results obtained. Finally, we also find essentially the same results when using different sample years, while controlling for the industry of the self-employed has a larger effect. The latter is not too surprising, though, as the choice of industry is closely correlated with the choice of education.

In sum, we therefore conclude that the preliminary OLS findings of substantial heterogeneity and non-linearities in the relationship between education and earnings seems to be a relatively robust result – at least in the Danish case. This has at least two implications. First, a methodological implication is that the log-linear Mincer specification is inappropriate in the case of entrepreneurs and self-employed. As we shall discuss, this conclusion is also supported by the fact that many self-employed experience negative earnings – something which cannot easily be handled in a Mincer framework. The second implication is that further research in this area is required to answer the following questions: Why are most types of education associated with very limited pecuniary pay-off for self-employed (as opposed to the case for wage employed)? Is it because education in itself is not sufficient for entrepreneurs, or is it because (the currently available type of) education is irrelevant for entrepreneurs? The answers to these questions may have substantial policy implications.

The rest of the paper is structured as follows: In Section 2, we present our empirical framework and discuss relevant identification issues. We describe the data used in Section 3. In Section 4, we present the main empirical results, while Section 5 contains results from a number of robustness checks.

Section 6 concludes.

2 The Empirical Framework

In this section, we outline the empirical framework and discuss a number of issues related to the estimation of the parameters of interest.

2.1 The General Specification

The general model is:

$$\ln Y_i = f(Edu_i) + \beta X_i + u_i \tag{1}$$

where Y_i is the earnings of individual i. $f(Edu_i)$ is a function of the educational attainment of the individual, Edu_i, which may be non-linear in years of schooling. X_i contains other characteristics of the individual, including experience, gender, region of residence *etc.*, and u_i is a random error.

In the standard Mincer equation (Mincer, 1974), $f(Edu_i)$ is simply a constant, α, times years of schooling:

$$f(Edu_i) = \alpha \cdot Years_i \tag{2}$$

Hence, a linear relationship between years of schooling and log-income is assumed. The coefficient, α, can in this case be interpreted as the percentage increase in market earnings from an extra year of schooling.

This specification is the starting point for the estimations in the present paper. However, we also consider a specification that includes both a linear and a quadratic term in years of schooling in order to analyze the importance of non-linearities. Moving on, we utilize our large and detailed dataset and consider different dummy variable specifications for the educational length. Finally, we interact the length dummies with dummies for different types of (higher) education. This last specification contains 22 dummy variables.

2.2 Estimation Issues

There are a number of data issues and econometric problems related to the estimation of (1). In this section, we provide a brief review of these problems and indicate how we deal with each of them in the present paper. The first problems are data related and concern how to define the entrepreneurs and identify their income. The remaining problems are econometric issues related to the estimation of (1), problems that potentially lead to inconsistent coefficient estimates.

First, there is the question of how to define/measure the entrepreneurs. In the present paper, we focus on the self-employed, *i.e.*, individuals owning an unincorporated business. It can be argued that owners of incorporated businesses are also entrepreneurs and hence should be included in the analysis. However, they are typically difficult to identify in the data, especially when we have register data as in the present case. The reason is that the owners of incorporated businesses are formally registered as employees when they work in their own firm and hence cannot be separated from the more "regular" wage workers. Furthermore, the observed wage and capital income of these persons need not be representative of the value generated by them as entrepreneurs if profits are saved in the firm. For these reasons, it is a common approach in the empirical entrepreneurship literature to focus on the self-employed, see also Parker (2004), although this may potentially create a selection bias if the relationship between log earnings and years of schooling is different for this class of entrepreneurs. As we explain below, we apply a Heckman correction model to deal with the non-random selection into self-employment vs. other occupations (including incorporated business owners). This should in principle solve the potential selection bias associated with excluding the owners of incorporated businesses from the group of entrepreneurs.

Focusing on the self-employed, there are at least two further data issues to be dealt with: *(i)* some individuals are both self-employed and wage workers at the same time; *(ii)* a number of individuals change status during the year, and given that we rely on annual observations, we have to determine whether these individuals should be included in the group of self-employed. To assure robustness of our results, we use three different definitions of the self-employed. Our preferred definition is the official definition of self-employment from Statistics Denmark. Each year at the last week of November, Statistics Denmark collects information regarding the primary occupation of each individual in Denmark. We rely on this information for our primary definition. As alternative (more narrow) definitions of the self-employed, we select: *(a)* the subset of self-employed with wage income below a certain threshold; and *(b)* the subset of self-employed with employees.

Using the first subset we remove individuals who have wage employment of any importance. Using the second removes many of those that have been self-employed for only part of the year since recently started self-employed are unlikely to have employees.

Second, there is the question of how to measure self-employment earnings. This problem stems from the fact that we typically have different measures of the reported income, and the fact that the reported income need not perfectly reflect the generated income. Hamilton (2000) uses three different measures of entrepreneurial returns, ranging from *net profit* to *equity adjusted draw*, where the latter is defined as the amount withdrawn for consumption plus the change in the equity of the company. In the present paper, we rely primarily on a measure of the *annual surplus* from self-employment activities which is very similar to the net profit measure used by Hamilton (2000). This amount is the one reported to the tax authorities and reflects the value added generated by the entrepreneur, and is typically different from the amount withdrawn for personal consumption. Following Hamilton, we assure robustness by using an additional earnings measure namely *gross annual income*, which is the total earnings of the entrepreneur including income from other sources.[3]

Turning to the econometric issues, these are relatively standard in the literature on returns to schooling and may result in non-zero correlation between the error term and the schooling measure thereby causing inconsistent estimates of the relevant parameters. The first problem relates to the fact that measures of education are likely to be endogenous as unobserved individual ability may affect both the choice of education and the earnings, as pointed out by Griliches (1977). In other settings with endogeneous right-hand-side variables, a possible response to this problem is to use a panel data set and estimate a fixed effects model. However, this is impossible in the present case as educational attainment is practically invariant across time for each individual, and hence the parameters of interest would not be identified in a fixed effects regression.

Instead, we try to deal with this problem by instrumenting the measure of education using two different sets of family background variables as in-

[3]None of these measures include non-pecuniary benefits. Hamilton (2000) argues that non-pecuniary benefits are likely to be important. But – like most other studies including Hamilton (2000) – we cannot control for this aspect. However, if these unmeasured benefits can be assumed to be proportional to the self-employment income, it will not bias the estimates of the returns to education, as we – as opposed to Hamilton (2000) – are not trying to compare returns to self-employment and wage work.

struments. More precisely we instrument years of schooling using the years of schooling of the spouse and the years of schooling of the parents. The instrumental-variables approach has been used in a number of studies in the traditional returns-to-schooling literature, but much less frequently in the literature which focuses on returns to schooling in self-employment and entrepreneurship.[4] Thus, to our knowledge only van der Sluis *et al.* (2008), Parker and van Praag (2006) and van der Sluis *et al.* (2007) try to deal with the endogeneity of schooling for entrepreneurs using IV techniques. In the traditional returns-to-schooling literature, the use of family background variables as instruments has been criticized; see, *e.g.*, Trostel *et al.* (2002). However, a few recent studies, such as Block, Hoogerheide, and Thurik (2010), indicate that these problems may not be that severe. Hence, we follow this approach in the present paper.

A second econometric issue is that the sample of self-employed does not necessarily constitute a random draw from the population. Instead, it consists of those who deliberately chose (or were "forced") to become self-employed – possibly because self-employment was relatively more advantageous than wage work to these people. Thus, we may have a classical non-random sample problem due to self-selection. We try to deal with this in the estimations using a Heckit procedure where we first estimate a probit selection model explaining the choice of self-employment versus being either wage-employed or un-/non-employed. A similar approach has been used by van der Sluis et al. (2007) who controls for selection into self-employment in addition to instrumenting education in a panel data setting. Second, based on this model we calculate the inverse Mills ratio which we then include in the regression of the final Mincer equation.

3 Data

The data we use in this study come from the Integrated Data Base for Labor Market Research ("IDA") compiled by Statistics Denmark. It contains register data since 1980 for all individuals living in Denmark. The data provide detailed information on labor market performance, such as past and present occupation, earnings and experience, as well as a wide range of background

[4] A third way of dealing with the endogeneity of schooling is to include proxies for unobserved ability, such as IQ test scores *etc.* In the wage-employment literature this has been tried by Griliches (1977).

characteristics like educational background and family characteristics.

Most of our analysis is conducted on a cross-section of self-employed from 2002. However, we exploit the panel structure of the dataset to construct a number of control variables. For instance, the labor market experience variables employed in the regressions and many of the family background variables which we use in the instrumental-variables estimations and in the Heckman selection models, are generated using the historical information in the data.

Job occupations in a given year are categorized according to an individual's primary labor market status in the last week of November. We are thus unable to control for flows between labor market states within a year. As mentioned above, our preferred definition of self-employed includes individuals who are characterized as being primarily self-employed by Statistics Denmark in 2002. However, as explained above, we also use two subsamples of these in order to exclude individuals with substantial wage employment "on the side" and individuals who have only been self-employed part of the year. The first subsample thus excludes individuals with wage income above DKK 25,000 ($\approx$ USD 5,600).[5] The threshold of DKK 25,000 was chosen because a relatively large group of self-employed have a tiny bit of wage income. In fact, about 7.9% of the self-employed in our sample have strictly positive wage income less than DKK 25,000, while a threshold of DKK 50,000 would only expand the sample by another 2.4%. The second subsample includes only self-employed with employees, as these are likely to have been self-employed the entire year.

As explained above, we use two income measures: *annual surplus* from self-employment activities and *gross annual income*. The cross section from 2002 consists of approximately 150,000 observations. Of these, approximately 20,000 observations have non-positive income. These are eliminated in order to be able to use log of income as the dependent variable. Throwing away 20,000 observations may create a selection bias. One potential solution would be to arbitrarily change the dependent variable to a small positive amount in the cases where non-positive income are observed. While this may solve the selection problem, it induces another problem in the subsequent regressions namely an imposed (one-way) measurement error on these observations,

[5]DKK 25,000 corresponds to about DKK 30,000 in 2010 which equals approximately \$5,600 (using the exchange rate on the 13th of November). Using the exchange rate in the last week of November 2002, the dollar amount is about \$3,000.

which may be quite significant in some cases.

Instead, we rely on the Heckit correction (explained above) to deal with this problem. In the Heckit procedure, the selection equation models the selection into the sample, *i.e.*, “self-employed with positive income” among all persons in the labor force. In order to specifically investigate the importance of the selection bias caused by throwing away the 20,000 observations with non-positive income, we also apply the Heckit correction to a more limited population. That is, we model the selection into the sample among all self-employed only.

Turning to the explanatory variables, in the literature either "years of schooling" or college and/or high-school dummies have been used to capture educational attainment; see van der Sluis *et al.* (2008). In the present paper, we rely on both a measure of schooling in years, as well as a large number of dummies for educational length and type. As additional control variables, we use a range of socio-demographic variables including experience in the labor market.

The Danish educational system includes a high variety of formal educational programmes, including vocational programmes as well as short, medium and long further educational programmes. A long further education corresponds to the Ph.D. or the master level (18+ years of total education). A medium further education corresponds to the bachelor level (16 years), whereas a short further education (14 years) is a shorter and more practical education than a bachelor degree. Primary and lower secondary school corresponds to nine and 10 years where nine years is the mandatory level in Denmark. A high school degree corresponds to 12 years. A vocational education is a mix of schooling and training in firms. The typical duration is three years, and results in a total of 12 years of education. Both the high-school and the vocational educational programmes are managed by the public sector which sets the standards and requirements for these types of education. This means that the quality and content of the various programmes are harmonized across schools assuring that individuals with a particular type of education have achieved training of comparable quality.

In the estimations, we operate with 22 different combinations of length (9, 10, 12, 14, 16, and 18+ years) and type (non-qualifying degree; humanities; natural science; social science; technical science; medical science and military). Table 1 contains summary statistics for the educational dummies for length and type for the 2002 cross section. It shows that 12 years of schooling is the most common ($> 50\%$) among the self-employed, and 80%

of those with 12 years of schooling are educated within a “technical” subject. Only around 10% have 18+ years of schooling, while a little less than 20% have the minimum length of schooling (9 years). The table also includes the dependent variable, log(earnings), and other background characteristics used in the estimations below. Most of these variables are self-explanatory. Note, however, that the experience variables, *self-employment experience* and *wage-employment experience* measure years of previous experience in wage- and self-employment since 1980 (the first year of the data), respectively. These variables are measures of actual labor market experience, and not just potential experience as typically used in the literature, where potential experience is calculated as a residual from the age of the individual and the length of his/her education; see Card (2001). *Spouse employed* is a dummy variable that takes the value one if the spouse assists in the firm, as this is likely to increase the annual surplus because the remuneration for this work is not (fully) deducted in the surplus. As in most of the literature on self-employment, we exclude farmers from the estimations.

< Insert Table 1 about here >

4 Empirical Results

As explained in the previous section, the results below are based on a cross-section of observations from 2002. For some of the robustness analyses in Section 5, we use cross sections from different years.

4.1 The Standard Specification

The first column in Table 2 contains the results of an OLS estimation of the "standard" Mincer equation where education is measured in years. This regression is similar to the one shown in Iversen *et al.* (2010) and included here as a benchmark. The difference between the OLS regressions of this paper and those in Iversen *et al.* (2010) is that we include regional dummies in the present paper. We observe that an extra year of schooling is in this case expected to yield an increase in earnings of 6.7%, see Table 1, Column

1.[6] This result is roughly in line with the existing literature. van der Sluis *et al.* (2008) thus report an average return of 6.1% across the 94 studies contained in their meta-analysis. The return to education estimated using the linear specification is illustrated in Figure 1 by the straight line, where the slope equals the point estimate of α.[7]

< Insert Figure 1 about here >

It should be noted that the estimated coefficients to the other control variables are all (highly) significant and generally of the expected sign. The estimated coefficients to the experience variables give us a crude indication of the importance of previous labor-market experience. The effect of self-employment experience is initially much larger than that of education but has the expected diminishing effect as experience accumulates since the coefficient to the square of self-employment experience is negative.[8] Wage-work experience also has a positive albeit smaller effect on income.

As noted above, a major problem in using OLS is that the measure of educational attainment, Edu_i, is likely to be endogenous in equation (1). If this is the case, the estimate of the return to schooling is both biased and inconsistent. To correct for this, Columns 2, 3 and 4 of Table 2 contain instrument-variable (IV) estimates using various sets of instrumental variables. Specifically, we consider three different sets: (i) the years of schooling of the parents, resulting in two instruments; (ii) the years of schooling of the spouse (one instrument); and (iii) the union of the first two sets. We use these instruments in a two-stage-least-squares estimation instrumenting the years of schooling of the self-employed. In the cases of the first and third instrument sets, the first-stage regression is overidentified whereas it is just identified when using the second set.

[6] To be precise, the estimate of α measures the log point change in income from an extra year of education. However, when the point estimate is small, this is approximately equal to the percentage change in income from an extra year of education.

[7] Actually, the slope is not completely constant as the percentage change in earnings from an extra year of schooling is only approximately constant (*cf.* the previous footnote).

[8] Note that self-employment experience is likely to capture other effects than human-capital accumulation. If self-employed invest in their firms when they are young and disinvest later on, this will create a positive correlation between the measured annual surplus and self-employment experience, which is due to physical- rather than human-capital accumulation.

The use of IV changes the estimated return to schooling. When we use the years of schooling of the parents as instruments (Column 2), the return to another year of education decreases to 3.7%. In contrast, when we use the years of schooling of the spouse as an instrument (Column 3), we obtain a return to education of 8.9% which is higher than the corresponding OLS estimate. The estimated return to education when we use years of schooling of both the parents and the spouse as instruments (Column 4) is 8.5%, close to the estimate obtained using only the education of the spouse as an instrument.

$<$ Insert Table 2 about here $>$

Table 2 also contain the F-statistics used for testing whether the instruments enter significantly in the first-stage regression, *i.e.*, a test of weak instruments. The statistics indicate that the instruments are highly significant in explaining the educational attainment of the self-employed. This is the case both when the education of the parents are used as instruments (Column 2), when the education of the spouse is used as an instrument (Column 3) and when the union of these are used (Column 4). In the first case, the F-test has a value of $2,641$, whereas in the second case, where only one instrument is used, the value is $21,375$, and in the last case, the value is $3,272$. In all cases, the magnitudes by far exceed the critical values of Stock *et al.* (2002) and Stock and Yogo (2005). Hence, a weak-instruments problem does not seem to be present.

Even though the instruments do not seem to be weak, there is still the question of whether they are valid, *i.e.*, uncorrelated with the error term in (1). It is possible to test this hypothesis when there are more instruments than endogenous variables, *i.e.*, when the first-stage regression is overidentified as in Columns 2 and 4. In the case where the years of schooling of the parents are used as instruments (Column 2), this test for overidentifying restrictions yields a p-value of 0.26.[9] Hence, we are unable to reject the null that our instruments are valid in this case. However, in Column 4, where all three instruments are used, the null is rejected with a p-value of 0.034.

[9] We report the Sargan score χ^2-test (Sargan, 1959). The alternative Basmann χ^2-test yields very similar results.

Given that the test cannot reject the validity of the first instrument set but rejects the validity of the third instrument set (where the education of the spouse is used in addition to the education of the parents), it is tempting to conclude that the years of schooling of the spouse is the only invalid instrument, and hence that the "correct" IV estimate of the return to another year of education is 3.7%. It should be noted, however, that the test we use might have low power for detecting endogeneity of the instruments (see Wooldridge, 2002). Hence, we cannot be certain that years of schooling of the parents are valid instruments. However, we can be fairly certain that at least some of our instruments are invalid, and the evidence seems to suggest that at least the years of schooling of the spouse is an invalid instrument.

The previous discussion suggests that the most plausible IV estimate of the return to schooling is lower than the corresponding OLS estimate. This would also be expected in the presence of unobserved individual ability which is positively correlated with both earnings and educational attainment. In contrast to this, previous studies on returns to schooling in self-employment have found IV estimates that are typically larger than the corresponding OLS estimates, see Parker and van Praag (2006) and van der Sluis *et al.* (2007). These studies also use various family background variables as instruments for education. For example, van der Sluis *et al.* (2007) use whether magazines were present in household at age 14, whether a library card was present in the household at age 14, the presence of a stepparent in the household and the number of siblings in the household as instruments for education. Parker and van Praag (2006) use an instrument very similar to ours as they instrument years of schooling of the self-employed with years of schooling of the father and the number of siblings in the respondent's family. Still, they find a larger effect of education when using IV instead of OLS.

The use of family background variables as instruments has been criticized in the wage-employment literature as unobserved individual ability is likely to be positively correlated across members of the same family. If this is the case, parental education will be correlated with subsequent earnings of the children since both are (partly) determined by family ability. Hence, parental education does not fulfill the exclusion restriction preventing its use as an instrument; see, *e.g.*, Trostel *et al.* (2002). However, a recent paper by Block, Hoogerheide, and Thurik (2010) indicates that the econometric problems related to the use of family background variables as instruments may not be that severe. Using a Bayesian approach, this paper concludes that the IV estimation results are robust to relaxing the exact validity assumption of the instruments.

Notice that the IV samples contain fewer observations than the OLS sample. For instance, when we use the years of schooling of the parents as instruments, we have a sample of about 50,000 individuals compared to the OLS sample of more than 130,000 individuals. This is because in order to use the schooling level of the parents, we need to establish a link between the self-employed person and his or her parents. For some of the individuals in our dataset, this link cannot be established. This is in particular the case for older individuals, which implies that the sample used for the IV estimation with years of schooling of the parents as instruments consists of younger individuals than the OLS sample. Also the sample used for the IV estimation where the instrument is the education of the spouse is smaller than the OLS sample. The former sample consists only of married (or cohabiting) self-employed which may differ from other self-employed in various ways. The bottom line of this discussion is that by using IV, we may introduce a sample-selection problem, since the IV samples do not consist of individuals that are drawn randomly from the entire group of self-employed.

To test if the differences between the OLS and IV estimates are, in fact, driven by the different samples rather than the instrumentation of "years of schooling", we also ran the OLS regressions on the reduced IV samples.[10] For the sample using years of schooling of the spouse as an instrument, we found an OLS estimate of 7.3%. This is slightly higher than the estimate of 6.7% on the full sample and goes some way towards explaining why the IV estimate on the same sample is 8.9%. In contrast, the OLS estimate on the sample using years of schooling of the parents as instruments is practically identical to the OLS estimate on the full sample: 6.6% vs. 6.7%. Hence, sample selection cannot explain why the IV estimate in this case drops to 3.7%.

As argued in Section 2.2, it is also possible that people self select into different occupations based on unobserved factors. Hence, the self-employed are possibly different from other members of the labor force in unobserved ways. If we want the estimated return to schooling to apply to the entire labor force, and not just the group of self-employed, we have to correct for this selection effect. We do this in the fifth column of Table 2 which contains estimates obtained using the Heckit procedure from Heckman (1979). Here, the probability of being in the sample is first estimated using a probit model, and

[10] The results are available upon request.

then the estimated inverse Mills ratio from this selection model is included as an extra regressor in the final Mincer equation (see, *e.g.*, Wooldridge, 2002).

Although the probit model can, in principle, be estimated using the same set of explanatory variables as in the final Mincer equation, identification of the parameters in the latter is typically weak if the same set of regressors is used in both models (Wooldridge, 2002). Hence, we use the number of years that the father and the mother have been self-employed since 1980 as extra regressors in the probit, as several studies have shown that children of self-employed parents are more likely to become self-employed themselves; see, *e.g.,* Hout and Rosen (2000) and Dunn and Holtz-Eakin (2000). The extra regressors are exclude them from the final regression. For this "exclusion restriction" to be valid, the number of years that the parents have been self-employed must not affect earnings conditional on being self-employed. This requires, *e.g.*, that self-employed with self-employed parents are no less likely to make "rookie" mistakes than self-employed without self-employed parents; an assumption which can always be questioned.

Table 2 shows that there is evidence of a sample-selection problem, since the estimated coefficient to the inverse Mills ratio is significant. Notice, however, that the estimated return to schooling of 7.6% is rather similar to the OLS estimate, so in this case correcting for sample selection does not change the conclusion drastically. There is still a significant and positive return to education for the self-employed.[11]

As explained in Section 3, we also apply the Heckit correction to a more limited population to investigate the importance of dropping the 20,000 observations with non-positive earnings. Obviously, self-employed with non-positive earnings might be different from those with positive earnings in unobserved ways inducing a potential sample-selection problem. We therefore ran a Heckit correction procedure where the underlying full sample consists of all self-employed (positive and negative earnings). The results are contained in the last column of Table 2. As for the other Heckit procedure, we find a significant inverse Mills ratio indicating that the sample is non-random. Still, we find a return to schooling which is very similar to the original estimate.

To sum up we find a return to an additional year of schooling of 6.7% in the standard OLS specification. The quantitative size of the return to

[11] Note that the dummy for the spouse assisting in the firm is not included in the selection model explaining self-employment as it does not seem reasonable to explain the choice of self-employment by the subsequent decision of the spouse to assist. Inlcuding it anyway does not change the results, however.

schooling changes somewhat when estimated by IV methods, where we find both smaller and larger returns to schooling compared to the OLS case. In what seems to be the least problematic IV estimation, the estimate is only 3.7%. We also find returns to education of similar magnitude as in the OLS case when we correct for non-random samples using Heckman's approach. Hence, a general conclusion from this section is that education seems to carry positive returns in self-employment.

4.2 A Quadratic Specification

The main point of Iversen *et al.* (2010) is that the returns to schooling in self-employment seem to be both non-linear in years of schooling and heterogenous across different types of education. In this section, we investigate the first point in more detail by including a quadratic term in years of schooling, Edu_i^2, in the Mincer regression. Our estimates are presented in Table 3 below.

The first column contains OLS results similar to those from Iversen *et al.* (2010) and are included here for completeness. It is seen that the coefficients to both the linear and the quadratic term in years of schooling become highly significant, while all other parameters remain largely unaffected compared to the standard specification, except for the constant term (*cf.* Table 2). To illustrate the non-linearity, Figure 1 includes the return profile based on the OLS estimates in the first column of Table 3. The estimated return to schooling implies a strongly increasing marginal return to education, but, as it turns out, 12 (or less) years of schooling provide no extra return compared to nine years of schooling (the mandatory level).

To test the robustness of these results, we again consider both instrumental variables regressions and use Heckit correction procedures. With respect to the IV regressions, we use the same sets of instrumental variables as in the previous section. Hence, we run three IV regression using the years of schooling of the parents, the years of schooling of the spouse and the union of these sets as instruments. Of course, we now have two potentially endogenous regressors, the linear and the quadratic term in years of schooling of the self-employed. We instrument these variables using both the linear and the squared values of the instruments (see Angrist and Pischke, 2010). This results in four, two and six instruments, respectively.

The IV results are presented in Columns 2-4 of Table 3. It is evident that the point estimates are very different from the results obtained under

OLS and that the point estimates depend on the instrumental variables used. For instance, when we use years of schooling of the parents as instruments (Column 2), the estimates suggest a concave relationship between years of schooling and earnings instead of the convex relationship found when using OLS. In the two other specifications, the importance of the quadratic term is found to be very small, suggesting an almost linear return profile. However, none of the IV estimates are significantly different from zero at the 5% level. Hence, it is difficult to draw strong conclusions about the shape of the relationship from the IV estimates.

As in the previous section, we are unlikely to suffer from a weak-instruments problem. However, the tests for overidentifying resitrictions in Columns 2 and 4 (the test is not applicable in Column 3) suggest that (some of) our instruments are invalid because they correlate with the error term in the Mincer equation. This should also warn us from drawing too strong conclusions based on the IV estimates in Table 3.

As an extra check, we also estimated an IV regression where only the linear versions of the instruments were used to explain both the linear and the quadratic term in years of schooling in the first stage (Column 5) In this case, we have two potentially endogenous variables and three instruments and therefore one overidentifying restriction. In this regression, the test of overidentifying restrictions cannot reject the null that the instruments are valid. Moreover, the coefficients to both the linear and the quadratic term in years of schooling are significant at the 5% level, and the point estimates support the convex relationship found in the OLS regression, although the point estimates differ somewhat. The estimated coefficient to the linear term is thus -1.78, while the estimated coefficient to the quadratic term is 0.07. These estimates imply a return profile where self-employed with 10-17 years of schooling actually have lower earnings than those with 9 years of schooling and where only 18 years of schooling yields a positive return compared to the mandatory level: self-employed with 18 years of schooling earn 41% more than than those with nine years of schooling. However, the standard errors associated with these IV estimates are also considerably larger than in the OLS case.

< Insert Table 3 about here >

We also ran Heckman correction models to account for the potential sample selection problems discussed previously. The results from these regressions are shown in the last two columns of Table 3. As for the linear models presented in Table 2, the significance of the estimated coefficients to the inverse Mills ratio indicate that the samples are indeed not randomly drawn. For the Heckit correction model where the population considered is the total labor force (Column 6), we get essentially the same conclusion as when using OLS: Earnings are a convex function of years of schooling and the estimated coefficients are of a similar magnitude. The same conclusion is reached when the population considered consists only of all the self-employed (Column 7).

To sum up, the convex relationship between years of schooling and log earnings generated using OLS is robust to corrections for non-random samples using the Heckit procedure. When using IV, the results are more mixed, but typically insignificant, and there are several indications that instruments are invalid. The only regression that pass the overidentification test and results in significant estimates that supports the convex relationship from the OLS estimation is when the linear versions of the instruments were used to explain both the linear and the quadratic term in years of schooling in the first stage. Hence, it seems fair to conclude that the evidence of significant non-linearities in the returns to schooling is robust to these extensions.

4.3 A Dummy Specification - Educational Length

As in Iversen *et al.* (2010), we also consider a specification which is fully flexible in years of schooling by including dummies for the different levels of schooling in the regression. The first column in Table 4 contains the OLS estimates of this regression and the results are similar to those from Iversen *et al.* (2010).

The reference category is nine years of schooling. Hence, the coefficient to *Dummy, 10 years of schooling* is the log point change in earnings from choosing 10 instead of nine years of schooling, which is approximately equal to the percentage change when the estimated coefficient is small. In what follows, we will convert an estimated coefficient into the exact percentage changes in earnings using the formula $(y_t - y_9)/y_9 = \exp(\beta_i) - 1$ where y_t indicates earnings of a self-employed with t years of schooling.

< Insert Table 4 about here >

According to the OLS results, three extra years of schooling result in 8.04% higher income ($\beta_{12} = 0.0773$), whereas further increases to 14 or 16 years of schooling have only minor effects. The return to seven years of extra schooling is thus only 9.32%, corresponding to an average return per year of schooling of approximately 1.2%. Moving from nine to 18 years of schooling, however, implies an increase in income of 117% ($\beta_{18} = 0.7763$), which is an average return of 9% per year of schooling. The coefficient estimates to the other control variables are similar to those in Table 2.

The OLS estimates are illustrated in Figure 1 by the dots. It is seen that only 18 years of schooling is associated with an increase in earnings compared to nine years of schooling which is economically significant, and this outlier is apparently driving the results of both the linear and the quadratic specifications.

We now consider the robustness of these findings. With respect to the IV approach, predicting five dummies for years of schooling using dummies for the educational attainment of the parents and the spouse turned out to be infeasible. Standard errors exploded in the final regressions resulting in insignificant or implausible estimates of the coefficients to all the relevant variables. In general, the price we must pay for using IV to get consistent estimators of the return to education is large confidence intervals. This was also the in the sections above, however, under the dummy specification this price is so high that we cannot pin down the coefficients with a reasonable precision.[12]

Hence, we focus on the results from using the Heckit correction procedure in what follows. Column 2 of Table 4 contains estimation results when the underlying population is the total labor force. The estimates support a highly non-linear relationship between years of schooling and log earnings. However, using the Heckit procedure, the return to 16 years of schooling is larger than in the OLS case. In the Heckit case, 16 years of schooling are associated with approximately 20% higher earnings compared to a self-employed with just nine years of schooling. In the OLS case, the difference was only about 10%. The coefficient estimates for the remaining variables all have the same sign and are in general of the same magnitude as in the OLS case.

As in the previous sections we also considered a Heckit model where the population of interest is all the self-employed and where the selection is into positive earnings. We still see a significant non-linear return to schooling

[12] The results are available upon request.

with the longest education programmes yielding quite substantial returns.

This point is clearly illustrated in Figure 2 which shows the point estimates of the three regressions in Table 4 converted into percentage changes in income relative to an individual with nine years of schooling. Hence, also in this case is it fair to conclude that the robustness analysis supports the OLS findings.

< Insert Figure 2 about here >

4.4 A Dummy Specification - Educational Type

Finally, we consider the full specification from Iversen *et al.* (2010), where (some of) the dummies for years of schooling are split into different types of education, resulting in 22 educational categories (plus the omitted category: 9 years of schooling). The OLS point estimates for this specification are contained in the first column of Table 5 and illustrated in Figure 3 below.[13] The figure shows very different effects – even conditional on the length of the education. The largest effects are obtained for 18 years of schooling within medical science (doctors and dentists), while 18 years of schooling within social science (including many lawyers and psychologists) also yields a substantial increase in earnings. Other types of long further education seem to carry very small or even negative returns. In other words, the returns to education within a given educational length are very heterogeneous for the self-employed. This is also true for 16 years of schooling, where substantial returns are found only within medical science.

< Insert Figure 3 about here >

To check the robustness of these findings, we also estimated this specification using the Heckit procedure to correct for non-random samples. The results from this are contained in the final two columns of Table 5. This did not change the general picture. The only estimates that are sensitive to

[13] In the table, we have excluded coefficient estimates on the control variables to save space. These results are available upon request.

changing specifications are for those educational types that are determined by relatively few observations and thereby become less precisely estimated. This is for example the case for 18 years of education within military, natural sciences and humanities. In general, the main result of heterogeneity of the returns to education is supported.

< Insert Table 5 about here >

In sum, this section has shown that the non-linearities and heterogeneity found in Iversen *et al.* (2010) using simple OLS is largely robust to extensions using IV methods to correct for the potential endogeneity of the schooling variables, and Heckit procedures to correct for the non-random samples. In the following Section, we consider other robustness checks.

5 Robustness Checks

The purpose of this section is to analyze the robustness of the main result of non-linearities and heterogeneity in the returns to education in self-employment that was established in the previous section. To perform the robustness analysis, we first study non-linearities separately using the specification from Section 4.3, after which we study the robustness of heterogeneity within each length of education, using the specification from Section 4.4. In particular, we want to investigate if the main result is sensitive to *(i)* the choice of sample year; *(ii)* the applied definition of self-employment; *(iii)* the applied measure of income; and *(iv)* other issues.

5.1 Non-Linearities

In this section, we study the robustness of non-linearities in the returns to education; a result that was illustrated in Figure 1 and presented in Table 4, Column 1. First, we study the importance of the choice of sample year by comparing the preferred OLS regression – that is based on sample year 2002 – to similar regressions for each of the sample years 1995, 1998, and 2001. The results are illustrated in Figure 4.

< Insert Figure 4 about here >

There are only minor differences over sample years between point estimates for the five dummy variables representing different years of schooling. This implies that the estimated returns to different years of schooling are insensitive to the choice of sample year. In other words, the choice of sample year does not influence the result of non-linearities. For the estimates we refer to Table A.1 in the appendix.

Second, we turn to the applied definition of self-employment. As discussed in Section 2.2, a potential problem is that the estimates are distorted as a consequence of our definition of self-employment. Statistics Denmark characterizes an individual as being self-employed if the main occupation is self-employment in the last week of November. The question is if we face an omitted variable bias since we do not include a covariate measuring how much of the year individuals have been self-employed. To deal with this issue, we use two alternative definitions of self-employment: *(a)* self-employed with earnings in wage employment of less than DKK $25,000$ in the sample year; and *(b)* self-employed with employees. The samples for both of these definitions are expected to consist of full-time self-employed to a higher extent than under the preferred specification. The regression results are illustrated in Figure 5 together with the results based on the preferred definition.

< Insert Figure 5 about here >

There are hardly any differences in point estimates across the different definitions of self-employment. This implies that the estimated returns to different years of schooling are insensitive to the definition of self-employment. In other words, the definition of self-employment does not influence the result of non-linearities. For the exact point estimates we refer to Columns 2–3 of Table A.2 in the appendix.

It is interesting to observe that the estimates in Columns 2–3 for alternative definitions of self-employment are very similar to the preferred specification. This is reassuring given that we expect the alternative definitions to capture the full-time self-employed most precisely. It is particularly interesting in the case of self-employed with employees since the sample size is reduced to 40 percent of the orginal sample size in this case. Using this definition, we are likely to introduce an additional selection problem since these individuals are the most successful self-employed. Still, the resulting estimates are not very different from those from our preferred specification.

Third, we study the robustness of non-linearities to the applied measure of income. As noted by Hamilton (2000), it is very difficult to accurately measure the income of self-employed. Given this difficulty, we want to investigate if the main result is sensitive to the use of an alternative income measure. In particular, we apply *gross annual income*. The results based on this income measure are illustrated in Figure 5, whereas the point estimates are presented in Column 4 of Table A.2.

It is seen that the point estimates are somewhat different from those of the preferred specification. For example, 16 years of schooling has a return of 23% under the alternative income measure compared to 9% for the preferred specification. Even though the exact returns differ between the two income measures, the point estimates generate the same picture of non-linearity as for the preferred specification.

Fourth, we investigate the robustness of non-linearity when the sample is restricted to include males only, to include individuals younger than 50 only, and by including a large set of industry dummies. The results are illustrated in Figure 6.

< Insert Figure 6 about here >

The differences between specifications are minor. The only estimate that is not fully robust is the return to 18 years of schooling in the specification that includes industry dummies. The finding that the returns to education fall significantly when industry dummies are included come as no surprise, as they may be strongly correlated with education itself. This implies that the inclusion of industry dummies will tend to lower the estimated return to education. This is precisely what we find for 18 years of schooling. The point estimates are presented in Tables A2, Columns 5–7.

5.2 Heterogeneity

To investigate the robustness of the pronounced heterogeneity in the returns to education within different years of schooling, we estimate the specification from Section 4.4 with dummy variables for years of schooling split into different types of education using similar changes as above for the robustness analysis of non-linearities. The estimates are available in Tables A3 and A4 in the appendix.[14]

The overall impression is that the returns to 18 years of education within medical science and social sciences are high as was the case in the above Figure 2. Moreover, the return to 18 years of education within technical sciences is also independent of the precise specification. On the other hand, the returns to 18 years of education within especially military and natural sciences change to a higher extend. This is due to relatively few observations within these educational types and thereby less precise estimates. For less than 18 years of schooling, a substantial return is only found within medical science with 16 years of schooling.

In sum, this section has shown that the main result of non-linearities and heterogeneity in the returns to education is robust to (among others) alternative choices of sample year, alternative definitions of self-employment, and alternative measures of income.

6 Conclusion

In this paper, we have studied the returns to education in self-employment based on very detailed register data on the Danish population. We use four different specifications to describe the relationship between log income and education and estimate these specifications using different estimation methods. For all specifications, we use OLS as well as Heckit correction models to handle sample selection. For the two most restricted specifications - see below - we also use two-stage-least-square estimation to handle endogenous regressors. The main result of the analysis is a high degree of non-linearity and heterogeneity in the returns to education in self-employment.

The first specification is the standard Mincer equation that specifies that an additional year of schooling is associated with a constant percentage increase in income. The conclusion from this analysis is that education carries a positive return in self-employment. Second, we include a quadratic term in years of schooling in the Mincer regression. Doing this we estimate a convex relationship between years of schooling and log income, implying that the

[14] In the tables, we have excluded coefficient estimates on the control variables to save space. These results are available upon request.

return is negligible for self-employed with 12 years of schooling compared to 9 years of schooling. Only for additional years of schooling can economically significant returns to education be detected. On this background, it is concluded that the relationship between log income and years of schooling is highly convex.

Next, we take the flexibility of the specification a step further and include dummies for the different years of schooling in the regression. We find that the return to education is highly non-linear in the educational length. Finally, we estimate the most flexible specification with dummy variables for years of schooling split into different types of education. We find that the returns to different types of education for a given educational length are heterogeneous. Our estimations indicate that the large returns are concentrated among certain educational types and that many educations hardly carry any return.

An immediate methodological implication of these findings is that the log-linear Mincer specification is inappropriate in the case of self-employed. This conclusion is further strengthened by another observation, namely that many self-employed have negative earnings. As we have discussed, this is not an issue that can be easily handled in the Mincer framework, where the dependent variable is the logarithm of earnings. However, while the non-linearities can be captured in a dummy specification, the problem with negative earnings requires an alternative approach, and the Heckman correction does not seem to be the most promising way of dealing with this. Hence, this is an obvious question to address in future research.

So why do people choose to educate themselves if the returns to education are so modest? It should be remembered that the estimations in this paper capture only one of the potential returns to education, namely the economic return that people obtain if they choose to become self-employed upon completing their education. As the vast majority of individuals end up in wage employment or move back and forth between wage employment and self-employment over their career, the total *expected* economic return when initiating an education is still likely to be significant; and to this could be added the non-economic returns to education such as self-esteem, a higher social status and being more knowledgeable.

Thus, our findings do not bear any immediate consequences for education policy – education is still likely to be a profitable investment in human capital. However, it may have important implications for entrepreneurship policy. The public debate about the importance of entrepreneurship has increased over the last couple of years and much of the literature identifies entrepreneurs with self-employed. It is generally argued that entrepreneurship fosters growth both in production and employment and is therefore desirable from the viewpoint of society. This obviously raises the question of which factors determine whether an entrepreneur becomes successful. One candidate is obviously education. The estimations in this paper show that some educations can raise income by as much as 230% (medical science 18+) compared to mandatory education. The estimations also show that for most educational programmes, however, the returns are miniscule and not comparable to those found in wage work. So what then makes a successful (rich) entrepreneur?

An obvious road for future research is to look for other elements of "entrepreneurial ability". Lazear (2004) have proposed a theory of entrepreneurs as "jacks of all trades", and Lucas (1978) and Malchow-Møller *et al.* (2010) have argued for the potential importance of skills acquired in previous wage work. Our results indicate that previous wage work experience but also previous self-employment experience are significant determinants of returns in self-employment.

References

[1] Angrist, J. D. and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton University Press, Princeton.

[2] Becker, G. (1964): *Human Capital*, National Bureau of Economic Research, New York.

[3] Block, J. H., L. Hoogerheide and R. Thurik (2010): "Are Education and Entrepreneurial Income Endogenous and do Family Background Variables make Sense as Instruments? A Bayesian Analysis", *Tinbergen Institute Discussion Paper*, TI 2010-024/4.

[4] Card, D. (1999): "The Causal Effect of Education on Earnings", in O. C. Ashenfelter and D. Card (ed.), *Handbook of Labor Economics*, vol 3(1), pp. 1801-1863, Elsevier, Amsterdam.

[5] Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, vol. 69(5), pp. 1127 - 1160.

[6] Chevalier, A., G. Conlon, F. Galindo-Rueda, and S. McNally (2002): *The Returns to Higher Education Teaching,* Research Report to the Department of Education and Skills, Centre for the Economics of Education, London.

[7] Dunn, T. and D. Holtz-Eakin (2000): "Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links", *Journal of Labor Economics*, vol. 18(2), pp. 282-305.

[8] Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, vol. 45(1), 1-22.

[9] Hamilton, B. H. (2000): "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment", *Journal of Political Economy*, vol. 108(3), 604-631.

[10] Harmon, C., H. Oosterbeek and I. Walker (2003): "The Returns to Education: Microeconomics", *Journal of Economic Surveys*, vol. 17(2), pp. 115 - 156.

[11] Heckman, J. J. (1976): "A Life Cycle Model of Earnings, Learning and Consumption", *Journal of Political Economy*, vol. 84(4), pp 11-44.

[12] Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, vol. 47(1), 153-161.

[13] Heckman, J. J., L. J. Lochner and P. E. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond", in E. Hanushek and F. Welch (ed.), *Handbook of the Economics of Education*, vol. 1, pp. 307 - 458, Elsevier, Amsterdam.

[14] Hout, M. and H. Rosen (2000): "Self-Employment, Family Background, and Race", *Journal of Human Resources*, vol. 35(4), pp. 670-692.

[15] Iversen, J., N. Malchow-Møller and A. Sørensen (forthcoming): "Returns to Schooling in Self-Employment", Economics Letters.

[16] Lazear, E. (1977): "Education: Consumption or Production?", *Journal of Political Economy*, vol. 85(3), pp 569-598.

[17] Lazear, E. P. (2004): "Balanced Skills and Entrepreneurship", *American Economic Review*, vol. 94(2), pp. 208 - 211.

[18] Lucas, R. E. Jr., (1979) : "On the Size Distribution of Business Firms", *Bell Journal of Economics*, vol. 9(2), pp. 508 - 523.

[19] Malchow-Møller, N. , J.R. Markusen, J.R. Skaksen. (2010) Labour market institutions, learning and self-employment. Small Business Economics 35:1, 35-52.

[20] Mincer, J. (1974): *Schooling, Experience and Earnings*, National Bureau of Economic Research, New York.

[21] Parker, S. C. (2004): *The Economics of Self-Employment and Entrepreneurship*, Cambridge University Press, Cambridge.

[22] Parker, S. C. and M. van Praag (2006): "Schooling, Capital Constraints and Entrepreneurial Performance: the Endogenous Triangle", *Journal of Business and Economic Statistics,* Vol. 24(4), pp. 416 - 431.

[23] Stock, J.H., M., Yogo and J.H. Wright, 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics, 20, 518-529.*

[24] Stock, J.H. and M. Yogo, 2005. Testing for Weak Instruments in Linear IV Regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.* In D.W.K. Andrews and J.H. Stock (eds), Cambridge University Press, pp. 80-108.

[25] Trostel, P., I. Walker and P. Woolley (2002): "Estimates of the economic return to schooling for 28 countries", *Labour Economics*, vol. 9(1), pp. 1 - 16.

[26] van der Sluis, J., M. van Praag, and A. van Witteloostuijn (2007): "Why are the Returns to Education Higher for Entrepreneurs than for Employees?", *IZA Discussion Paper*, No. 3058.

[27] 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*, vol. 22(5), pp. 795 - 841.

[28] Walker, I. and Y. Zhu (2001): *The Returns to Education: Evidence from the Labour Force Surveys,* Research Report 313, Department of Education and Skills, London.

[29] Wooldridge, J. M. (2002): *Econometric Analysis of Cross-Section and Panel Data,* MIT Press, Cambridge, MA.

Table 1: Summary Statistics

	Mean	Std. Dev.	No. Obs.	Min	Max
Log(earnings)	12.1437	1.2897	131447	0	19.9519
Yrs. of schooling	12.3182	2.6093	131,447	9	18
Dummy, 9 yrs. schooling	0.1827	-	131,447		
Dummy, 10 yrs. schooling	0.0673	-	131,447		
Dummy, 12 yrs. schooling	0.5261	-	131,447		
Dummy, 14 yrs. schooling	0.0456	-	131,447		
Dummy, 16 yrs. schooling	0.0803	-	131,447		
Dummy, 18 yrs. schooling	0.0980	-	131,447		
Dummy, non qualifying, 9 yrs.	0.1827	-	131,447		
Dummy, non qualifying, 10 yrs.	0.0673	-	131,447		
Dummy, non qualifying, 12 yrs.	0.0583	-	131,447		
Dummy, humanities, 12 yrs.	0.0523	-	131,447		
Dummy, humanities, 14 yrs.	0.0099	-	131,447		
Dummy, humanities, 16 yrs.	0.0249	-	131,447		
Dummy, humanities, 18 yrs.	0.0079	-	131,447		
Dummy, natural sciences, 16 yrs.	0.0005	-	131,447		
Dummy, natural sciences, 18 yrs.	0.0023	-	131,447		
Dummy, social sciences, 14 yrs.	0.0037	-	131,447		
Dummy, social sciences, 16 yrs.	0.0159	-	131,447		
Dummy, social sciences, 18 yrs.	0.0243	-	131,447		
Dummy, technical, 12 yrs.	0.4024	-	131,447		
Dummy, technical, 14 yrs.	0.0279	-	131,447		
Dummy, technical, 16 yrs.	0.0239	-	131,447		
Dummy, technical, 18 yrs.	0.0198	-	131,447		
Dummy, medical, 12 yrs.	0.0131	-	131,447		
Dummy, medical, 14 yrs.	0.0019	-	131,447		
Dummy, medical, 16 yrs.	0.0145	-	131,447		
Dummy, medical, 18 yrs.	0.0430	-	131,447		
Dummy, military, 14 yrs.	0.0022	-	131,447		
Dummy, military, 16 yrs.	0.0006	-	131,447		
Dummy, military, 18 yrs.	0.0007	-	131,447		
Age	48.1613	12.2203	131,447	15	87
Dummy, male	0.7292	-	131,447		
Dummy, married	0.7816	-	131,447		
Dummy, immigrant	0.0748	-	131,447		
Dummy, city	0.6649	-	131,447		
Self-employment experience	10.5758	7.4690	131,447	1	23
Wage-employment experience	8.4710	6.4032	131,447	0	22
Spouse employed	0.0409	-	131,447		

Note: Sample is 2002 cross-section and includes non-farm self-employed with positive earnings.

Table 2: Returns to schooling, linear specification, results from 2002 cross-section, dependent variable is log of annual surplus

	OLS	IV, Parents edu.	IV, Spouse edu.	IV, All Instru.	Heckit, labor force	Heckit, all self-emp.
Years of schooling	0.0666	0.0374	0.0893	0.0849	0.0760	0.0631
	*53.10****	*5.31****	*26.87****	*16.08****	*34.63****	*21.83****
Age	0.0692	0.0541	0.0670	-0.0113	0.0054	0.0046
	*32.32****	*7.16****	*25.05****	*-1.16*	*0.81*	*0.5*
Age, squared	-0.0010	-0.0008	-0.0010	-0.0001	-0.0003	-0.0003
	*-46.98****	*-8.42****	*-39.29****	*-0.61*	*-3.80****	*-2.78***
Dummy, male	0.3117	0.2853	0.4324	0.3824	0.2042	0.1460
	*42.07****	*24.85****	*51.36****	*30.56****	*12.72****	*4.36****
Dummy, married	0.0880	0.1697	-	-	0.1399	0.0863
	*11.04****	*14.22****	-	-	*12.10****	*3.98****
Dummy, immigrant	-0.0488	0.0660	-0.0670	0.0470	-0.0069	0.1773
	*-3.55****	*1.55*	*-4.13****	*0.84*	*-0.21*	*4.25****
Dummy, city	0.1033	0.0400	0.0943	0.0388	0.0497	0.0458
	*12.55****	*3.10****	*10.40****	*2.81***	*4.13****	*2.89***
Self-employment experience	0.1233	0.1613	0.1279	0.1754	0.1572	0.1586
	*49.75****	*41.64****	*44.02****	*39.50****	*44.98****	*39.86****
Self-employment experience, squared	-0.0017	-0.0031	-0.0016	-0.0037	-0.0028	-0.0028
	*-17.12****	*-17.18****	*-14.17****	*-17.46****	*-17.12****	*-15.21****
Wage-employment experience	0.0662	0.1045	0.0660	0.0800	0.0969	0.0956
	*26.18****	*22.94****	*21.82****	*14.58****	*23.62****	*20.63****
Wage-employment experience, squared	-0.0007	-0.0016	-0.0007	-0.0005	-0.0013	-0.0012
	*-6.10****	*-8.33****	*-4.76****	*-2.28**	*-7.24****	*-6.03****
Spouse employed	0.6554	0.4965	0.6552	0.5065	0.5219	0.5247
	*39.87****	*13.76****	*40.21****	*14.41****	*16.25****	*13.81****
Inverse Mills ratio	-	-	-	-	-0.2901	-1.4832
	-	-	-	-	-7.60***	*-4.95****
Constant	8.3271	8.6129	8.1804	9.4772	9.8778	10.0211
	*170.43****	*56.41****	*124.59****	*51.24****	*51.37****	*32.29****
Regional-dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	131,447	49,598	100,878	36,992	56,913 1,747,885	56,913 64,687
R^2	0.2009	0.2061	0.2131	0.1926	-	-
F-statistic, IV	-	2,640.76***	21,374.63***	3,272.39***	-	-
Chi^2-statistic, IV	-	1.2580	-	6.7344*	-	-

Note: *t* or *z* statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level. The OLS sample consists of all non-farm self-employed with positive earnings. The sample for IV (parents) is the OLS sample excluding individuals without parental education information. The IV (spouse) sample is OLS sample excluding individuals without a spouse. The IV (all) sample is the intersection of the IV (parents) and IV (spouse) samples. The uncensored observations in Heckit (labor force) are all non-farm self-employed with parental occupation information available. The censored observations are all other in the labor force. The uncensored observations in Heckit (self-emp.) are the same as in Heckit (labor force). The censored observations are non-farm self-employed with zero or negative earnings. The F-statistic refers to test of H0: coefficients on instruments are zero in the first stage regression. The Chi2-statistic refers to test of H0: no correlation between instruments and error term in second stage regression. The top (bottom) sample size for Heckit refers to number of uncensored (total) observations.

Table 3: Returns to schooling, quadratic specification, results from 2002 cross-section, dependent variable is log of annual surplus

	OLS	IV, Parents edu.	IV, Spouse edu.	IV, All Instru.	IV, All Instru. linear first stage	Heckit, labor force	Heckit, all self-emp.
Years of schooling	-0.2440	0.3238	0.0974	0.0793	-1.7877	-0.3045	-0.3343
	*-22.29****	*1.93*	*1.47*	*0.68*	*-2.27**	*-17.70****	*-13.72****
Years of schooling, squared	0.0116	-0.0103	-0.0003	0.0002	0.0676	0.0141	0.0148
	*28.56****	*-1.69*	*-0.12*	*0.06*	*2.38**	*22.21****	*16.61****
Age	0.0667	0.0563	0.0671	-0.0114	-0.0411	0.0000	-0.0013
	*31.20****	*7.29****	*24.81****	*-1.16*	*-2.53**	*0.00*	*-0.13*
Age, squared	-0.0010	-0.0008	-0.0010	-0.0001	0.0001	-0.0003	-0.0003
	*-46.63****	*-8.38****	*-39.28****	*-0.61*	*0.74*	*-3.53****	*-2.46**
Dummy, male	0.3139	0.2901	0.4323	0.3827	0.3434	0.2015	0.1403
	*42.49****	*24.62****	*51.35****	*29.85****	*16.24****	*12.58****	*4.07****
Dummy, married	0.0905	0.1659	-	-	-	0.1420	0.0864
	*11.38****	*13.64****	-	-	-	*12.33****	*3.87****
Dummy, immigrant	-0.0619	0.1118	-0.0667	0.0469	-0.2521	-0.0747	0.1094
	*-4.51****	*2.24**	*-4.07****	*0.79*	*-1.81*	*-2.30**	*2.56**
Dummy, city	0.1058	0.0352	0.0942	0.0385	0.0629	0.0508	0.0470
	*12.89****	*2.67***	*10.26****	*2.76***	*3.53****	*4.23****	*2.87***
Self-employment experience	0.1242	0.1606	0.1279	0.1755	0.1844	0.1593	0.1607
	*50.27****	*40.73****	*43.98****	*39.41****	*30.43****	*45.74****	*39.14****
Self-employment experience, squared	-0.0016	-0.0032	-0.0016	-0.0037	-0.0031	-0.0026	-0.0027
	*-16.73****	*-16.73****	*-14.07****	*-17.19****	*-9.37****	*-16.42****	*-14.10****
Wage-employment experience	0.0686	0.0978	0.0659	0.0800	0.1167	0.1037	0.1023
	*27.19****	*16.45****	*21.17****	*13.37****	*7.07****	*25.32****	*21.32****
Wage-employment experience, squared	-0.0007	-0.0015	-0.0007	-0.0005	-0.0012	-0.0014	-0.0013
	*-6.26****	*-6.85****	*-4.74****	*-2.21**	*-3.20***	*-7.92****	*-6.37****
Spouse employed	0.6652	0.4993	0.6550	0.5068	0.4894	0.5252	0.5282
	*40.58****	*13.74****	*40.12****	*14.41****	*12.87****	*16.42****	*13.47****
Inverse Mills ratio	-	-	-	-	-	-0.2942	-1.5300
	-	-	-	-	-	*-7.73****	*-4.94****
Constant	10.3647	6.6913	8.1260	9.5122	22.3734	12.4338	12.7126
	*119.99****	*5.91****	*18.14****	*11.55****	*4.12****	*54.17****	*32.98****
Regional-dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	131,447	49,598	100,878	36,992	36,992	56,913 1,747,885	56,913 64,687
R^2	0.2058	0.2044	0.2128	0.1928	0.0883	-	-
F-statistic, IV	-	1,369.77*** 1,425.70***	11,089.33*** 11,759.51***	1,756.93*** 1,896.84***	3,272.39*** 3,432.75***	-	-
Chi^2-statistic, IV	-	27.40***	-	19.34***	0.3201	-	-

Note: *t* or *z* statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level. The OLS sample consists of all non-farm self-employed with positive earnings. The sample for IV (parents) is the OLS sample excluding individuals without parental education information. The IV (spouse) sample is OLS sample excluding individuals without a spouse. The IV (all) sample is the intersection of the IV (parents) and IV (spouse) samples. The uncensored observations in Heckit (labor force) are all non-farm self-employed with parental occupation information available. The censored observations are all other in the labor force. The uncensored observations in Heckit (self-emp.) are the same as in Heckit (labor force). The censored observations are non-farm self-employed with zero or negative earnings. The F-statistic refers to test of H0: coefficients on instruments are zero in the first stage regression. The Chi2-statistic refers to test of H0: no correlation between instruments and error term in second stage regression. The top (bottom) sample size for Heckit refers to number of uncensored (total) observations

Table 4: Returns to schooling, dummies for years of schooling, results from 2002 cross-section, dependent variable is log of annual surplus

	OLS	Heckit, labor force	Heckit, all self-emp.
Dummy, 10 yrs. schooling	-0.0020	-0.0169	-0.0127
	-0.14	*-0.86*	*-0.47*
Dummy, 12 yrs. schooling	0.0773	0.0349	-0.0169
	*8.77****	*2.38**	*-0.75*
Dummy, 14 yrs. schooling	0.0832	0.0836	0.0507
	*5.00****	*3.41***	*1.54*
Dummy, 16 yrs. schooling	0.0891	0.1842	0.1041
	*6.53****	*7.39****	*3.53****
Dummy, 18 yrs. schooling	0.7763	0.8064	0.6909
	*60.68****	*37.75****	*19.98****
Age	0.0679	0.0042	0.0034
	*31.85****	*0.64*	*0.37*
Age, squared	-0.0010	-0.0003	-0.0003
	*-46.93****	*-4.11****	*-2.93****
Dummy, male	0.3093	0.2037	0.1426
	*41.94****	*13.07****	*4.31****
Dummy, married	0.0872	0.1401	0.0859
	*11.00****	*12.24****	*3.96****
Dummy, immigrant	-0.0537	-0.0741	0.1090
	*-3.92****	*-2.27**	*2.58***
Dummy, city	0.1041	0.0501	0.0464
	*12.73****	*4.19****	*2.84****
Self-employment experience	0.1221	0.1563	0.1577
	*49.55****	*44.96****	*38.36****
Self-employment experience, squared	-0.0017	-0.0026	-0.0027
	*-17.00****	*-16.53****	*-14.15****
Wage-employment experience	0.0671	0.0994	0.0978
	*26.67****	*24.29****	*20.36****
Wage-employment experience, squared	-0.0007	-0.0013	-0.0012
	*-6.10****	*-7.53****	*-6.00****
Spouse employed	0.6670	0.5276	0.5308
	*40.82****	*16.55****	*13.53****
Inverse Mills ratio	-	-0.2989	-1.5302
	-	*-7.77****	*-5.07****
Constant	9.0662	10.7871	10.8149
	*187.95****	*54.10****	*35.76****
Regional-dummies	Yes	Yes	Yes
Sample Size	131,447	56,913 1,747,885	56,913 64,687
R^2	0.2109	-	-

Note: *t* or *z* statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level. The OLS sample consists of all non-farm self-employed with positive earnings. The uncensored observations in Heckit (labor force) are all non-farm self-employed with parental occupation information available. The censored observations are all other in the labor force. The uncensored observations in Heckit (self-emp.) are the same as in Heckit (labor force). The censored observations are non-farm self-employed with zero or negative earnings.

Table 5: Return to types of schooling, results from 2002 cross-section, dependent variable is log of annual surplus

	OLS	Heckit, labor force	Heckit, all self-emp.
Dummy, non qualifying, 10 yrs.	-0.0030	-0.0180	-0.0090
	-0.21	*-0.93*	*-0.25*
Dummy, non qualifying, 12 yrs.	0.0189	-0.0882	-0.0167
	1.23	*-4.29****	*-0.41*
Dummy, humanities, 12 yrs.	-0.0673	-0.2395	-0.3418
	*-4.13****	*-5.45****	*-4.01****
Dummy, humanities, 14 yrs.	-0.0713	-0.2053	-0.1381
	*-2.18**	*-4.09****	*-1.49*
Dummy, humanities, 16 yrs.	-0.2186	-0.0162	-0.0708
	*-10.21****	*-0.43*	*-1.11*
Dummy, humanities, 18 yrs.	0.0705	0.2514	0.1042
	1.95	*4.65****	*1.03*
Dummy, natural sciences, 16 yrs.	-0.9895	-0.8682	-0.8636
	*-6.84****	*-6.06****	*-3.27***
Dummy, natural sciences, 18 yrs.	-0.0684	0.2077	0.0622
	-1.05	*2.17**	*0.35*
Dummy, social sciences, 14 yrs.	0.0407	0.1635	0.2763
	0.78	*2.92***	*2.60***
Dummy, social sciences, 16 yrs.	0.1853	0.0330	0.1421
	*7.12****	*0.82*	*1.80*
Dummy, social sciences, 18 yrs.	0.7215	0.7315	0.6064
	*33.39****	*21.25****	*9.35****
Dummy, technical, 12 yrs.	0.1127	0.0709	-0.0094
	*12.39****	*4.72****	*-0.31*
Dummy, technical, 14 yrs.	0.1562	0.1437	0.0747
	*7.71****	*5.01****	*1.37*
Dummy, technical, 16 yrs.	0.1289	0.2678	0.2020
	*5.93****	*6.98****	*2.95***
Dummy, technical, 18 yrs.	0.3497	0.4830	0.4008
	*14.85****	*12.78****	*5.68****
Dummy, medical, 12 yrs.	0.0428	0.1799	-0.0018
	1.48	*3.98****	*-0.02*
Dummy, medical, 14 yrs.	0.1668	0.2119	-0.0696
	*2.33**	*1.88*	*-0.32*
Dummy, medical, 16 yrs.	0.4701	0.5023	0.1681
	*17.14****	*13.82****	*1.99**
Dummy, medical, 18 yrs.	1.2018	1.0741	0.9462
	*70.47****	*31.55****	*12.89****
Dummy, military, 14 yrs.	-0.1652	-0.2722	-0.5332
	*-2.47**	*-2.18**	*-2.26**
Dummy, military, 16 yrs.	-0.2395	-1.4178	-1.2144
	-1.86	*-3.77****	*-1.79*
Dummy, military, 18 yrs.	-0.3639	0.5386	0.4329
	*-3.08***	*1.67*	*0.72*
Inverse Mills ratio	-	-0.3337	-2.0990
	-	*-8.59****	*-5.88****
Additional Controls	Yes	Yes	Yes
Sample Size	131,447	56,913 1,747,885	56,913 64,687
R^2	0.2259	-	-

Note: *t* or *z* statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level. The OLS sample consists of all non-farm self-employed with positive earnings. The uncensored observations in Heckit (labor force) are all non-farm self-employed with parental occupation information available. The censored observations are all other in the labor force. The uncensored observations in Heckit (self-emp.) are the same as in Heckit (labor force). The censored observations are non-farm self-employed with zero or negative earnings. Additional controls are the same as in

Table A.1: Return to schooling, dummies for years of schooling, different years, dependent variable is log of annual surplus

	Preferred Specification, 2002	1995	1998	2001
Dummy, 10 yrs. schooling	-0.0020	0.0453	0.0163	-0.0097
	-0.14	*2.99***	*1.11*	*-0.65*
Dummy, 12 yrs. schooling	0.0773	0.0676	0.0757	0.0590
	*8.77****	*8.08****	*8.92****	*6.49****
Dummy, 14 yrs. schooling	0.0832	0.0808	0.1559	0.0878
	*5.00****	*4.51****	*9.16****	*5.03****
Dummy, 16 yrs. schooling	0.0891	0.0815	0.1132	0.0700
	*6.53****	*5.65****	*8.14****	*4.92****
Dummy, 18 yrs. schooling	0.7763	0.7567	0.7611	0.7865
	*60.68****	*58.08****	*59.47****	*58.98****
Age	0.0679	0.0934	0.0930	0.0961
	*31.85****	*44.20****	*44.60****	*44.35****
Age, squared	-0.0010	-0.0013	-0.0013	-0.0013
	*-46.93****	*-57.90****	*-60.34****	*-61.49****
Dummy, male	0.3093	0.4251	0.3840	0.3262
	*41.94****	*55.29****	*51.32****	*42.35****
Dummy, married	0.0872	0.1115	0.1255	0.1402
	*11.00****	*13.55****	*15.67****	*17.06****
Dummy, immigrant	-0.0537	-0.0939	-0.0983	0.0094
	*-3.92****	*-6.08****	*-6.93****	*0.66*
Dummy, city	0.1041	0.0800	0.0815	0.0696
	*12.73****	*9.79****	*10.02****	*8.14****
Self-employment experience	0.1221	0.2289	0.1582	0.1540
	*49.55****	*57.69****	*50.51****	*56.75****
Self-employment experience, squared	-0.0017	-0.0059	-0.0024	-0.0024
	*-17.00****	*-26.55****	*-16.28****	*-20.88****
Wage-employment experience	0.0671	0.0609	0.0771	0.0692
	*26.67****	*15.78****	*24.49****	*24.98****
Wage-employment experience, squared	-0.0007	0.0012	-0.0004	-0.0002
	*-6.10****	*4.46****	*-2.25**	*-1.67*
Spouse employed	0.6670	0.7541	0.7317	0.7084
	*40.82****	*61.55****	*54.03****	*43.53****
Constant	9.0662	7.9105	8.2326	8.2092
	*187.95****	*166.93****	*175.60****	*168.76****
Regional-dummies	Yes	Yes	Yes	Yes
Sample Size	131,447	127,482	129,763	130,970
R^2	0.2109	0.2595	0.2526	0.2437

Note: All models estimated with OLS. t statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level.

Table A.2: Return to schooling, dummies for years of schooling, different sample specifications, dependent variable is log of annual surplus unless otherwise indicated

	Preferred Specification	Wage inc. <= DKK 25,000	Self-emp. w/employee	Gross Annual Income	Males only	Age <= 50	Industry dummies
Dummy, 10 yrs. schooling	-0.0020	-0.0135	-0.0304	0.0351	0.0314	-0.0313	0.0099
	-0.14	*-0.9*	*-1.58*	*3.61****	*1.84*	*-1.88*	*0.69*
Dummy, 12 yrs. schooling	0.0773	0.0771	0.0360	0.0615	0.1000	0.0525	0.1156
	*8.77****	*8.59****	*3.08***	*10.41****	*10.21****	*4.34****	*12.95****
Dummy, 14 yrs. schooling	0.0832	0.0795	0.0749	0.1025	0.1256	0.0678	0.0907
	*5.00****	*4.61****	*3.35***	*9.18****	*6.75****	*3.28***	*5.45****
Dummy, 16 yrs. schooling	0.0891	0.1033	0.1072	0.2302	0.1052	0.1077	0.0889
	*6.53****	*7.26****	*5.29****	*25.21****	*6.60****	*5.90****	*6.42****
Dummy, 18 yrs. schooling	0.7763	0.7919	0.7548	0.7452	0.7109	0.7956	0.5425
	*60.68****	*58.55****	*50.63****	*86.93****	*49.12****	*44.28****	*36.46****
Age	0.0679	0.0633	0.0475	-0.0113	0.0729	0.0651	0.0642
	*31.85****	*28.87****	*13.98****	*-7.92****	*29.79****	*11.07****	*30.32****
Age, squared	-0.0010	-0.0010	-0.0006	0.0001	-0.0011	-0.0010	-0.0009
	*-46.93****	*-44.80****	*-18.46****	*4.81****	*-45.72****	*-13.75****	*-44.26****
Dummy, male	0.3093	0.3200	0.2821	0.2699	-	0.3051	0.2430
	*41.94****	*42.02****	*28.78****	*54.61****	-	*33.77****	*30.13****
Dummy, married	0.0872	0.0863	0.0619	0.0379	0.2176	0.1364	0.0866
	*11.00****	*10.52****	*5.64****	*7.13****	*23.71****	*14.01****	*11.08****
Dummy, immigrant	-0.0537	-0.0601	-0.2027	-0.1373	0.0107	0.0725	-0.0595
	*-3.92****	*-4.27****	*-10.65****	*-14.95****	*0.68*	*4.39****	*-4.29****
Dummy, city	0.1041	0.1043	0.0385	0.0640	0.0821	0.0354	0.1053
	*12.73****	*12.34****	*3.92****	*11.68****	*8.91****	*3.39***	*13.00****
Self-employment experience	0.1221	0.1124	0.0961	0.0615	0.1203	0.1558	0.1157
	*49.55****	*43.38****	*25.91****	*37.25****	*40.44****	*51.80****	*47.24****
Self-employment experience, squared	-0.0017	-0.0013	-0.0015	-0.0005	-0.0010	-0.0027	-0.0017
	*-17.00****	*-12.65****	*-10.38****	*-8.01****	*-8.41****	*-20.05****	*-17.56****
Wage-employment experience	0.0671	0.0683	0.0409	0.0543	0.0840	0.0892	0.0561
	*26.67****	*26.08****	*10.29****	*32.21****	*27.51****	*27.84****	*22.51****
Wage-employment experience, squared	-0.0007	-0.0005	-0.0006	-0.0005	-0.0011	-0.0010	-0.0005
	*-6.10****	*-4.29****	*-3.22***	*-6.37****	*-8.01****	*-7.12****	*-4.02****
Spouse employed	0.6670	0.6635	0.3775	0.1282	0.5904	0.5361	0.6664
	*40.82****	*40.52****	*21.57****	*11.71****	*35.38****	*19.52****	*41.18****
Constant	9.0662	9.2753	10.2591	11.3109	9.0821	8.9632	9.2447
	*187.95****	*185.29****	*139.35****	*349.94****	*170.48****	*83.62****	*168.86****
Regional-dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	131,447	116,219	51,988	131,401	95,848	72,507	131,447
R^2	0.2109	0.2227	0.1941	0.1897	0.2245	0.2268	0.2353

Note: All models estimated with OLS. t statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level.

Table A.3: Return to types of schooling, different years, dependent variable is log of annual surplus	Preferred Specification	1995	1998	2001
Dummy, non qualifying, 10 yrs.	-0.0030	0.0388	0.0103	-0.0137
	-0.21	*2.58**	*0.71*	*-0.92*
Dummy, non qualifying, 12 yrs.	0.0189	-0.0477	-0.0515	-0.0762
	1.23	*-2.96***	*-3.38***	*-4.92****
Dummy, humanities, 12 yrs.	-0.0673	-0.1375	-0.1326	-0.0923
	*-4.13****	*-8.69****	*-8.32****	*-5.47****
Dummy, humanities, 14 yrs.	-0.0713	-0.0542	0.0742	-0.0699
	*-2.18**	*-1.52*	*2.18**	*-2.05**
Dummy, humanities, 16 yrs.	-0.2186	-0.3248	-0.2398	-0.2654
	*-10.21****	*-12.70****	*-10.31****	*-11.75****
Dummy, humanities, 18 yrs.	0.0705	0.0418	0.0534	-0.0253
	1.95	*0.90*	*1.29*	*-0.65*
Dummy, natural sciences, 16 yrs.	-0.9895	-0.8663	-0.5401	-1.5220
	*-6.84****	*-3.88****	*-2.91***	*-9.73****
Dummy, natural sciences, 18 yrs.	-0.0684	-0.0345	0.1437	0.0266
	-1.05	*-0.42*	*2.03**	*0.39*
Dummy, social sciences, 14 yrs.	0.0407	0.1242	0.1316	0.1700
	0.78	*1.27*	*1.98**	*2.95***
Dummy, social sciences, 16 yrs.	0.1853	0.1719	0.1676	0.1126
	*7.12****	*6.27****	*6.33****	*4.24****
Dummy, social sciences, 18 yrs.	0.7215	0.7140	0.6668	0.7313
	*33.39****	*30.57****	*29.48****	*31.91****
Dummy, technical, 12 yrs.	0.1127	0.1154	0.1310	0.1094
	*12.39****	*13.29****	*14.90****	*11.63****
Dummy, technical, 14 yrs.	0.1562	0.1451	0.2096	0.1554
	*7.71****	*6.86****	*10.30****	*7.32****
Dummy, technical, 16 yrs.	0.1289	0.1730	0.2323	0.1648
	*5.93****	*7.73****	*10.73****	*7.24****
Dummy, technical, 18 yrs.	0.3497	0.3751	0.3961	0.3746
	*14.85****	*15.45****	*16.98****	*15.35****
Dummy, medical, 12 yrs.	0.0428	0.0419	-0.0622	0.0121
	1.48	*1.38*	*-2.10**	*0.40*
Dummy, medical, 14 yrs.	0.1668	-0.1814	0.0165	0.1915
	*2.33**	*-2.13**	*0.22*	*2.52**
Dummy, medical, 16 yrs.	0.4701	0.4077	0.4096	0.4888
	*17.14****	*13.15*	*13.78****	*16.7****
Dummy, medical, 18 yrs.	1.2018	1.0868	1.1370	1.2157
	*70.47****	*63.19*	*66.78****	*68.53****
Dummy, military, 14 yrs.	-0.1652	-0.0649	-0.0748	-0.2654
	*-2.47**	*-0.80*	*-1.06*	*-3.87****
Dummy, military, 16 yrs.	-0.2395	-0.3300	-0.2141	-0.4213
	-1.86	*-2.46**	*-1.72*	*-3.31***
Dummy, military, 18 yrs.	-0.3639	-0.4827	-0.2836	-0.3667
	*-3.08***	*-3.95****	*-2.39**	*-3.12***
Additional Controls	Yes	Yes	Yes	Yes
Sample Size	131,447	127,482	129,763	130,970
R^2	0.2259	0.2709	0.2651	0.2588

Note: All models estimated with OLS. *t* statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level. Additional controls are the same as in Tables 1-4

Table A.4: Return to types of schooling, different sample specifications, dependent variable is log of annual surplus unless otherwise indicated

	Preferred Specification	Wage inc. <= DKK 25,000	Self-emp. w/employee	Gross Annual Income	Males only	Age <= 50	Industry dummies
Dummy, non qualifying, 10 yrs.	-0.0030	-0.0141	-0.0290	0.0362	0.0290	-0.0261	0.0085
	-0.21	*-0.95*	*-1.51*	*3.76****	*1.72*	*-1.58*	*0.59*
Dummy, non qualifying, 12 yrs.	0.0189	0.0211	0.1638	0.1389	0.0308	-0.0554	0.0365
	1.23	*1.32*	*7.81****	*13.58****	*1.77*	*-3.14***	*2.36**
Dummy, humanities, 12 yrs.	-0.0673	-0.0823	-0.0183	-0.1509	0.0890	0.0731	0.1517
	*-4.13****	*-5.03****	*-0.85*	*-13.84****	*3.21***	*3.39***	*6.89****
Dummy, humanities, 14 yrs.	-0.0713	-0.1020	0.1610	-0.0060	-0.0564	-0.0983	-0.0340
	*-2.18**	*-3.06***	*2.36**	*-0.27*	*-1.10*	*-2.38**	*-1.04*
Dummy, humanities, 16 yrs.	-0.2186	-0.2096	-0.0896	0.1019	-0.2370	-0.1088	-0.1622
	*-10.21****	*-9.20****	*-2.63***	*7.11****	*-8.41****	*-3.92****	*-7.56****
Dummy, humanities, 18 yrs.	0.0705	0.0586	0.2241	0.2564	0.0251	0.1574	0.0865
	1.95	*1.53*	*2.84***	*10.59****	*0.51*	*3.54****	*2.40**
Dummy, natural sciences, 16 yrs.	-0.9895	-1.1342	-	-0.3225	-0.8896	-0.9466	-0.9011
	*-6.84****	*-7.67****	-	*-3.33***	*-6.00****	*-7.02****	*-6.27****
Dummy, natural sciences, 18 yrs.	-0.0684	0.0042	-0.2285	0.3228	-0.0385	0.0182	-0.0259
	-1.05	*0.06*	*-2.12**	*7.37****	*-0.54*	*0.24*	*-0.40*
Dummy, social sciences, 14 yrs.	0.0407	0.0204	-0.0240	0.1578	0.0939	0.1106	0.0905
	0.78	*0.37*	*-0.33*	*4.50****	*1.58*	*2.16**	*1.74*
Dummy, social sciences, 16 yrs.	0.1853	0.2133	0.2000	0.3020	0.2289	0.0178	0.1955
	*7.12****	*7.69****	*5.36****	*17.34****	*8.22****	*0.51*	*7.49****
Dummy, social sciences, 18 yrs.	0.7215	0.6622	0.6172	0.6969	0.6676	0.7102	0.6860
	*33.39****	*27.91****	*23.20****	*48.18****	*27.39****	*23.92****	*31.04****
Dummy, technical, 12 yrs.	0.1127	0.1148	0.0312	0.0823	0.1127	0.0821	0.1338
	*12.39****	*12.40****	*2.61***	*13.51****	*11.33****	*6.58****	*14.47****
Dummy, technical, 14 yrs.	0.1562	0.1502	0.0894	0.1276	0.1680	0.1255	0.1502
	*7.71****	*7.18****	*3.55****	*9.42****	*8.02****	*4.96****	*7.37****
Dummy, technical, 16 yrs.	0.1289	0.1187	0.0701	0.2270	0.1651	0.1743	0.1264
	*5.93****	*5.23****	*2.07**	*15.61****	*7.47****	*5.50****	*5.80****
Dummy, technical, 18 yrs.	0.3497	0.3362	0.3640	0.4059	0.3719	0.4403	0.2645
	*14.85****	*13.46****	*11.09****	*25.75****	*15.04****	*13.35****	*10.93****
Dummy, medical, 12 yrs.	0.0428	0.0703	-0.1000	0.0611	0.1356	0.0938	0.0528
	1.48	*2.37**	*-2.11**	*3.16***	*1.38*	*2.63***	*1.84*
Dummy, medical, 14 yrs.	0.1668	0.2208	0.3066	0.2331	0.4674	0.2246	0.1458
	*2.33**	*2.97***	*3.07***	*4.86****	*3.41***	*2.73***	*2.06**
Dummy, medical, 16 yrs.	0.4701	0.4895	0.3554	0.3973	0.5703	0.4783	0.4086
	*17.14****	*17.39****	*8.53****	*21.64****	*11.10****	*15.00****	*14.70****
Dummy, medical, 18 yrs.	1.2018	1.2763	0.8898	1.0481	1.1106	1.2689	0.9389
	*70.47****	*70.52****	*53.74****	*91.82****	*55.62****	*51.70****	*39.45****
Dummy, military, 14 yrs.	-0.1652	-0.0416	-0.2839	0.0977	-0.1191	-0.1676	-0.1557
	*-2.47**	*-0.56*	*-3.40***	*2.19**	*-1.82*	*-2.13**	*-2.36**
Dummy, military, 16 yrs.	-0.2395	-0.1955	0.1897	0.0516	-0.1857	-0.3758	-0.2239
	-1.86	*-1.52*	*0.84*	*0.60*	*-1.48*	*-2.24**	*-1.76*
Dummy, military, 18 yrs.	-0.3639	-0.4932	0.3862	0.4954	-0.2886	0.0869	-0.3055
	*-3.08***	*-4.04****	*1.54*	*6.27****	*-2.53**	*0.39*	*-2.61***
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	131,447	116,219	51,988	131,401	95,848	72,507	131,447
R^2	0.2259	0.2400	0.2035	0.2069	0.2361	0.2408	0.2429

Note: All models estimated with OLS. t statistics are in italics. *,**,*** indicate significance at 5%, 1% and 0.1% level. Additional controls are the same as in Tables 1-4

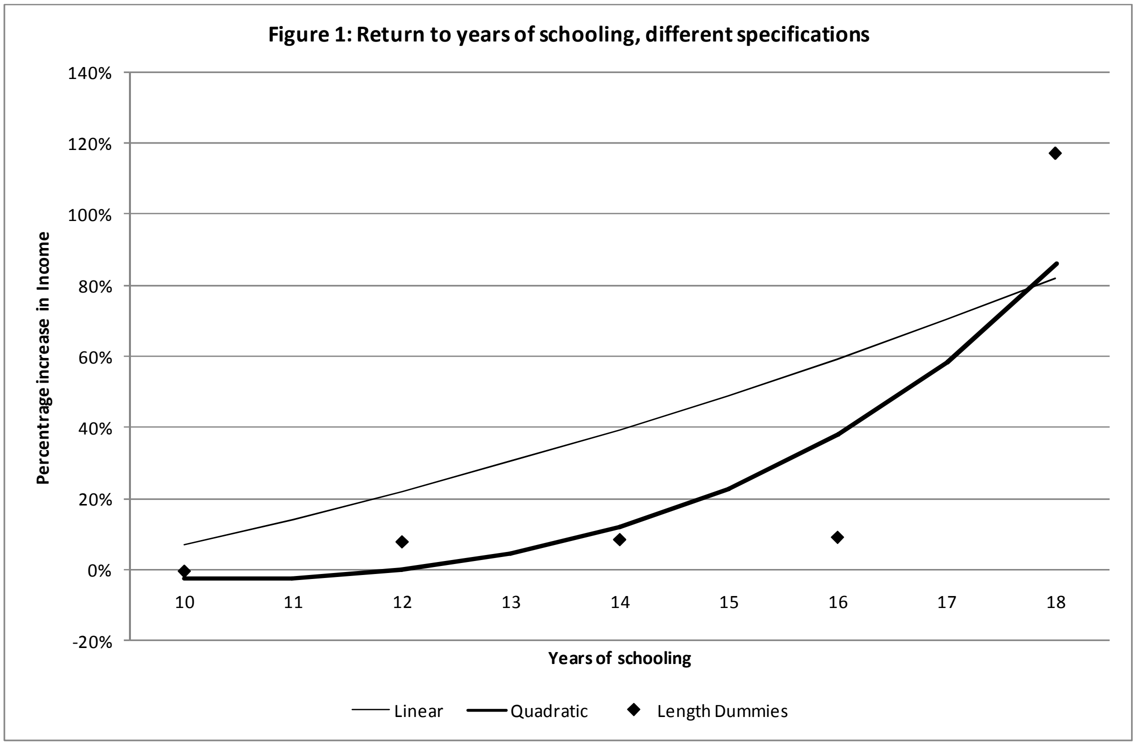


Figure 1: Return to years of schooling, different specifications
140%
120%
100%
80%
60%
40%
20%
0%
-20%
Percentage increase in Income
10
11
12
13
14
15
16
17
18
Years of schooling
Linear
Quadratic
Length Dummies

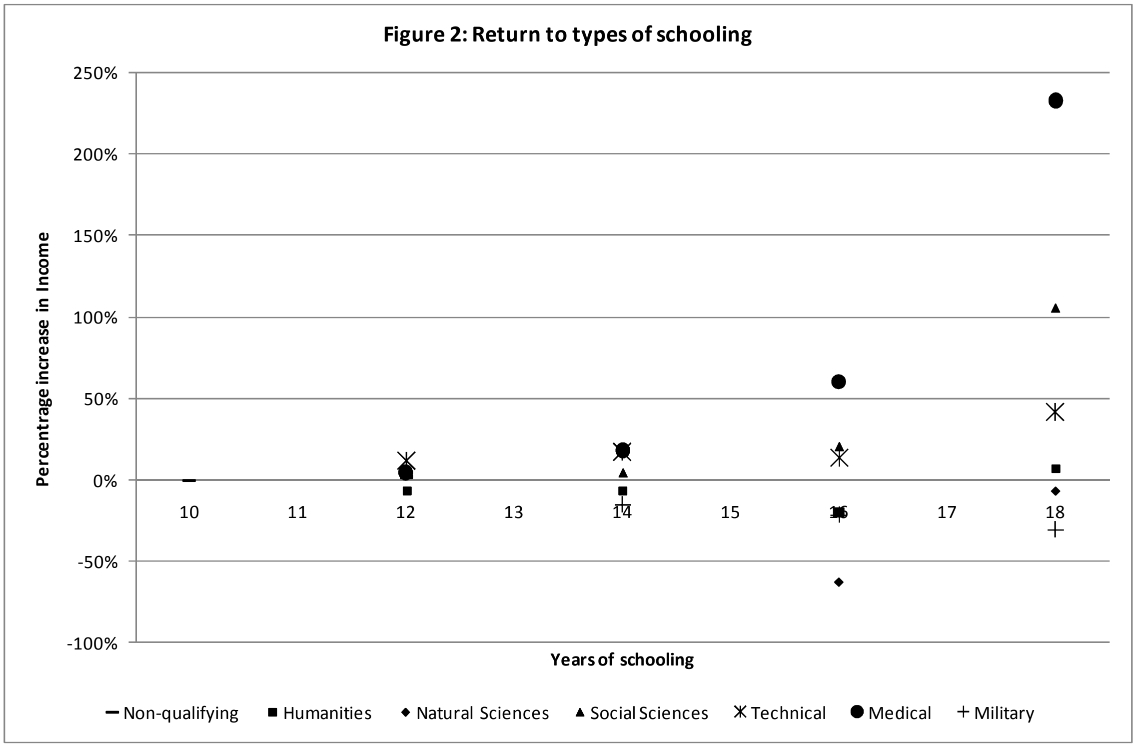
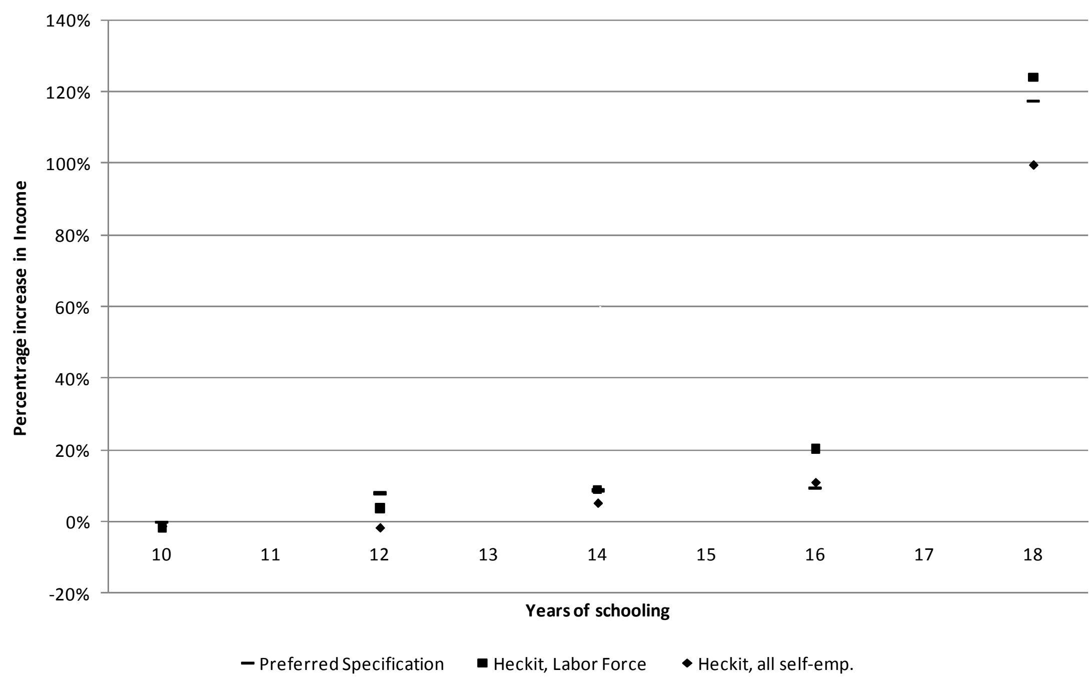


Figure 2: Return to types of schooling
250%
200%
150%
100%
50%
0%
-50%
-100%
Percentrage increase in Income
10
11
12
13
14
15
16
17
18
Years of schooling
Non-qualifying
Humanities
Natural Sciences
Social Sciences
Technical
Medical
Military

Figure 3: Return to years of schooling, length dummies, different specifications



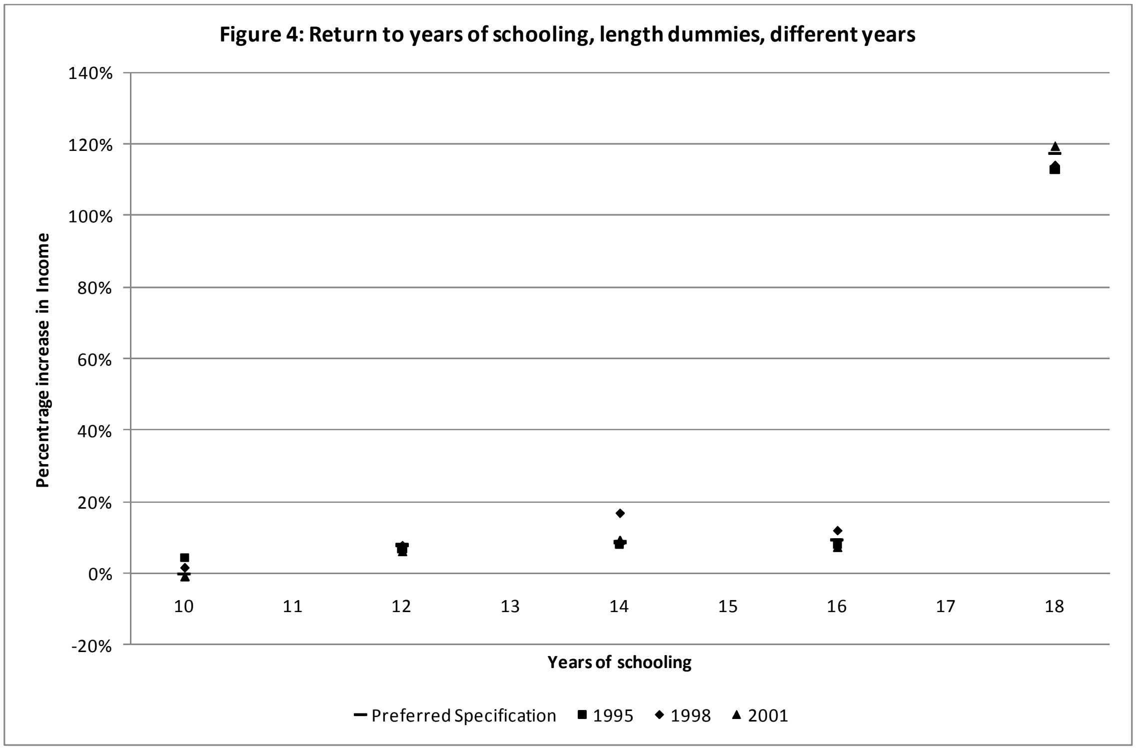
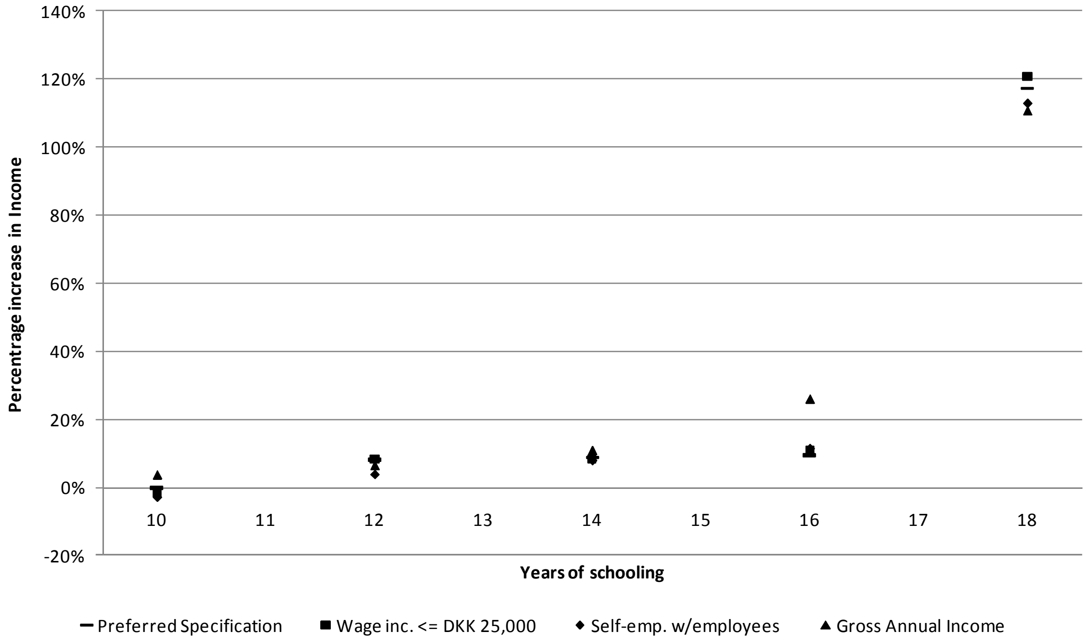


Figure 4: Return to years of schooling, length dummies, different years
140%
120%
100%
80%
60%
40%
20%
0%
-20%
Percentrage increase in Income
10
11
12
13
14
15
16
17
18
Years of schooling
Preferred Specification
1995
1998
2001

Figure 5: Return to years of schooling, length dummies, different definitions of sample and earnings



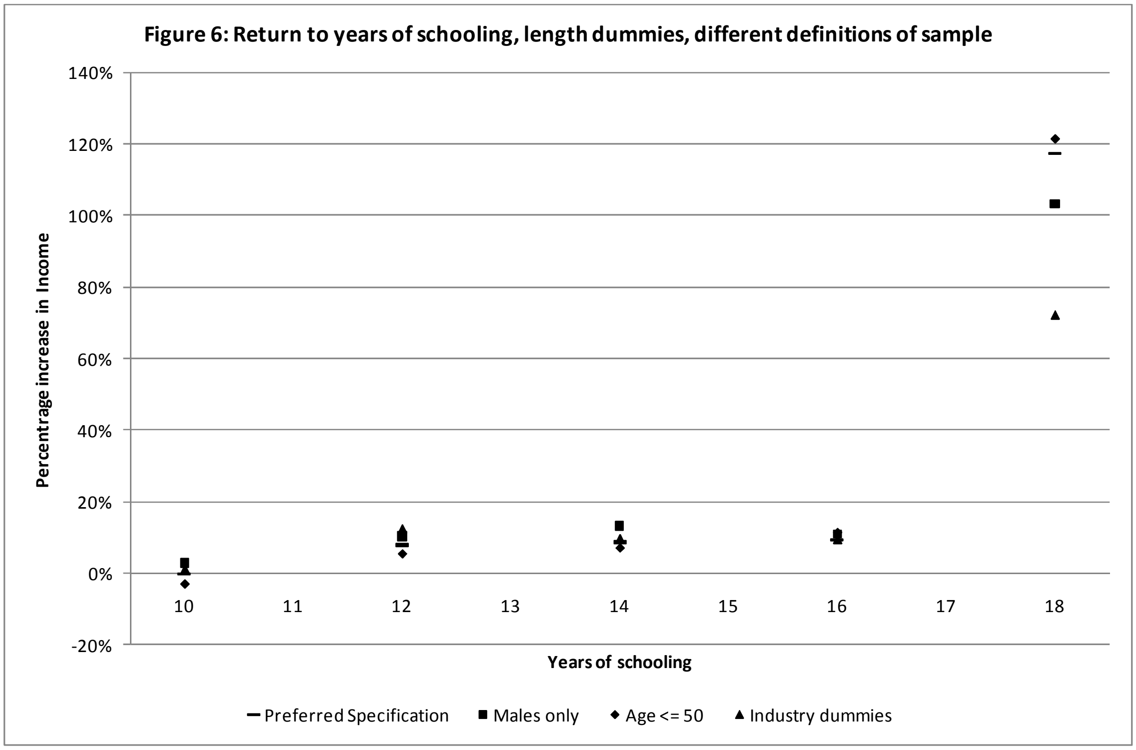


Figure 6: Return to years of schooling, length dummies, different definitions of sample
140%
120%
100%
80%
60%
40%
20%
0%
-20%
Percentrage increase in Income
10
11
12
13
14
15
16
17
18
Years of schooling
Preferred Specification
Males only
Age <= 50
Industry dummies