Returns to Education in Spain

Laura Requena
lrequenacasado1@sheffield.ac.uk

University of Sheffield, United Kingdom

April 10, 2015

Abstract

This paper estimates the returns to education in Spain using an instrumental variables approach. In order to overcome the endogeneity problem, a two stage least squares estimation technique is applied. The Spanish data come from the Living Conditions Survey in 2011. In this paper, only wage-earners are taken into account. This study reasserts the use of family background variables as instruments for schooling. Two different schooling variables are instrumented: years of schooling and levels of schooling. Several sets of instruments are tried. Depending on the level of schooling instrumented (primary, secondary first stage, secondary second stage and higher education) different sets of instrumental variables are chosen after trying combinations of instruments for each level. In agreement with the literature reviewed, ordinary least squares coefficients are downward biased. The substantive results of this exercise are entirely congruent with the theory of human capital: as expected, after using a set of variables to instrument the educational levels, individuals who attained higher education have the largest returns, although in Spain the returns to education obtained by women are higher than by men. Interestingly, father's education has a larger impact than mother's education on the years of education attained by their children and, therefore, on the returns to their education.

1 Introduction

A quote from Weisbrod (1962) states the importance of education: "Schooling benefits many persons other than the student. It benefits the student's future children who will receive informal education at home; it benefits neighbors who may be affected favorably by the social values developed in children. Schooling benefits employers seeking a trained labour force; and it benefits society at large by developing the basis for an informed electorate." Even if in this study we only focus on the effect of schooling in wages, education has more benefits than the increase on wages.

To continue assessing how important education is today, we quote some words by Andrea Schleicher, OECD Deputy-Director of Education: 'It is shown that those from disadvantaged backgrounds have still low chances to succeed in education; even in the richest nations in the OECD, the chances of moving up to high education are only about half if the parents didn't succeed in school'. It can be seen that those who didn't complete high school have seen rapidly deteriorating prospects'. Since the turn of the century, highly educated people have seen their job prospectives grow dramatically. The medium and low educated have seen theirs decline.

Why study the Returns to Education in Spain and not the causes of the economic growth in Ghana? After reviewing the literature, I can see that there is not an agreement about the exact returns to education.

Why is it still a challenge to find the definitive returns? Depending on the models that you use or the variables that you include, the coefficients change (sometimes dramatically). Spanish politicians are concerned about education. In fact, they believe that education is the engine that promotes the development of a country; if a citizen acquires more education, this will enable him/her to get to a higher position in the workplace and this will help the economic growth of the country. What the politicians are trying to avoid is an increase in the drop out rate and youth unemployment.

The exact coefficient for the private returns to education and what estimation method should be used is today still a topic of discussion. Researchers debate what variables should the model include as covariates in the wage equation and which estimation method is the one that performs best?

In this paper, the two stages least squares method is used to overcome the endogeneneity problem as a result of measurement error in the schooling variable. What it is empirically explored is which instrumental variables perform best to explain the difference in the attained education. The vast majority of papers from the review of the literature of the returns to education in Spain focus more on instrumenting years of schooling. In this study both education variables are instrumented: years and levels. By finding suitable instruments for the levels of education attained we can see what variables affect the attainment of the consecutive higher level of education.

This paper is divided as follows: first, a literature review regarding returns to education is given in section 2; section 3 explains the data set used for the analysis and the variables included in the models; section 4 describes the methodology employed in the estimation; section 5 reports the results obtained; section 6 concludes the paper.

2 Literature Review and Research Hypothesis

It is well known that there is a huge amount of literature on returns to education. Mincer's wage equation is one of the most used tools to obtain these returns. Some authors that have used this equation in their papers are: Psacharopoulos (1994) and Harmon, Oosterbeek, and Walker (2003). The first one is a paper comparing returns among different countries; estimation of returns by gender and sector of employment are computed. The second one explores functional forms, measures of schooling and social and private returns to education. In this study private returns to education are chosen to be estimated rather than social returns.

Researchers and politicians are interested in both returns. Education is important to single individuals because it allows them to earn more, have a better standard of living and belong to a higher social class. This also has a benefit for the Government in terms of taxes: if the individual earns more, he/she contribute more. The reason why private returns are estimated rather than social returns is because of the availability and accuracy of earnings.

Mincer (1974) proposed a semi-logarithmic wage equation to estimate the private returns to education. In this equation, wages vary in a linear way with the time invested in education and in a quadratic way with work experience:

$$\ln(W_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \varepsilon_i \tag{1}$$

W_i is a wage measure for individual i, S_i is the amount of schooling that individual i has, E_i is work experience, E_i^2 its square and ε_i represents the error term, that is assumed to be independent of the rest of variables in the equation and would capture other variables not included in the model. There are some assumptions in this model that we list: (i) there are no direct costs associated with the investment in education (only the opportunity cost of not receiving a salary because of studying), (ii) the individual stays constantly in the labour market (they do not leave the labour market); the time that they stay in the labour market is independent of the level of schooling attained and (iii) the individuals start to work right away after they finish their studies (no gap years).

The estimations of the returns to education have been controversial since Mincer proposed this equation. Critics emerged due to econometric problems that researchers can incur if they use Mincer's method. Griliches (1977) lists the problems that will cause the coefficients to be biased: (i) if there is existence of omitted variables such as ability; (ii) the non correct measure of education (measurement error) and (iii) education treated as an exogenous (not affected by anything inside the estimation equation) variable. The majority of the literature focuses on solving these econometric problems in order to get unbiased estimates. Some papers that stand out for doing this are Blackburn and Neumark (1993), Angrist and Krueger (1990, 1992) and Card (1999). Blackburn and Neumark (1993) in their paper explore whether the ordinary least squares (OLS) estimates of the returns to schooling are downwards bias. They study the omitted-ability bias. In order to do this, the National Longitudinal Survey Youth Cohort (NLSY) is used. Test scores are used to isolate the two ability components: academic and non-academic. They conclude that compared to OLS estimates, the education effect is lower when controlling for ability. If ability is not taken into account and instrumental variables are used to solve the endogeneity problem, the

returns to education are higher than the OLS ones.

Angrist and Krueger (1990) wrote a well known paper using data from United States. The returns to schooling are estimated using month of birth as an instrumental variable for education. Basically what these authors propose is that the quarter of birth when an individual was born makes a difference to the amount of education attained by that individual. The idea behind this is that a person born in the beginning of the year (first quarter) starts school later and reaches the minimum school leaving age before those born in later months, therefore these individuals would have less schooling than those born at the end of the year. There is a direct relationship then between the attained education and quarter of birth. Also, there is no reason to think that quarter of birth would have a direct effect on earnings. Several researchers question the validity of this paper. One of these is Bound, Jaeger, and Baker (1995). They find that the estimates that Angrist and Krueger report suffer from finite-sample bias so therefore are not consistent as expected. Another paper that criticized Angrist and Krueger's work is Hoogerheide and Dijk (2006) that states that the coefficient obtained in Angrist and Krueger is mainly determined by data from a few Southern states and also, they conclude that the instrument quarter of birth does not affect all the population with the same strength: the individuals who have 9-13 years of schooling are not that affected in comparison with the ones that have at most 8 or at least 14 years of schooling. How the instrument quarter of birth work is updated because the initial idea was that this instrument affects only the ones about leaving school (who normally would have about 9-13 years of schooling). Now they report that the instrument affects both groups, and specifically in a more strong way the second group.

Card (1999) reviews the literature on the causal effect of education on earnings. He focuses on four areas: theoretic and econometric progress, recent studies that exploit the institutional features of the education system to obtain the returns to education using instrumental variables, recent studies using samples of twins and new studies to evidence the presence of heterogeneity in the returns to education. What he concludes is that coefficients obtained in twins studies are biased upwards about 10 percentage points because of ability. In the studies where institutional changes in the education system are used, the bias between the OLS and the IV approaches ranges between 20 - 40 percentage points. The explanation that Card gives is that the marginal returns for certain groups (especially for those with low education levels) is larger than the average marginal returns in the full population.

The papers listed above are some of the "classic" ones in the economics of education literature. Due to the extensive literature related to this topic, this section is divided into two subsections, separating an international evidence and Spanish literature review concerning the returns to education.

There are two important pieces of work that review the Spanish literature review on this topic. The first one is written by Salas (2002); it reviews the origins of the Economics of Education and has a section that explains how this field proliferates in Spain. The second one is a recent meta-analysis done by Pino et al. (2014).

To review the studies that are already published about the returns to education in Spain the recent meta-analysis done by Pino et al. (2014) is first mentioned. It compiles more than 80 papers on returns to education in Spain. As these authors said, the first study about the returns to

education in Spain was published in 1978 and since then, every year there are new papers about this topic. To write this meta-analysis, several search engines and databases were used. The study encompasses papers written in English and Spanish.

Some of the general conclusions that the authors extract after doing the meta-analysis are the following: because of data availability, the majority of Spanish papers use a continuous measure for education (years rather than levels or degree attainment), the wage premium is larger for men than women and the coefficients obtained by using ordinary least squares are smaller than the ones obtained using the instrumental variables estimation method.

What we can conclude, after reviewing the empirical evidence from Spain, is that any researcher who wants to estimate the returns to education for this country is very restricted because of limited availability of data sets and lack of important variables. However, the list of papers reviewed before offer an initial idea of what is the rate of return to education in this country.

The way this paper fills the gap in the literature is by instrumenting dummy levels of education and by trying different sets of instrumental variables for each level to see which one adjusts best to explain specific levels of education.

According to Requena and Bernardi (2005), the growth of the education system in Spain in the last forty five years has been mainly due to two factors: education reforms and an increase of the education resources. Although the implementation of the reforms was gradual and not performed at an equal pace in all the regions, thanks to them eventually an expansion in the education system took place in Spain. Today, the economic system has grown and education, particularly higher education, has become more accessible to all the people, no matter their socioeconomic status. Moreover, the persistent fall of fertility rates in Spain in the last decades has made the number of students smaller and therefore, as public investments in education increased over time, the expenditure in education per student in now higher than before. Some basic characteristics that define modernization of the Spanish education system during the 20th century are: (1) the almost complete disappearance of illiterate (there are now only a few old people who do not know how to write or read); (2) the total elimination of child labour (almost all the recent generations have received education from 4 to 16 years); (3) the growing gender equality in terms of allowing women to go further in education; and (4) an easier access to higher education. A convenient way to summarize the heart of this issue is to say that 80 percent of Spaniards of all the cohorts born after 1960 have reached, at least, the first stage of secondary education.

Therefore, taking into account all the changes in the education system and the intense process of social modernization that Spain has experienced in the second half of the last century, the main hypotheses of this paper are formulated as follows: since (a) at the macroeconomic level, the rise in the proportion of people acquiring secondary education was promoted by growing public investment in education, and because (b) at the microeconomic level, traditional Spanish mothers spent more time with their children than fathers did (which is explained with more detail in the results section), it can be expected that public expenditure in education and mother's education will be valid instruments for the attainment of lowest levels of education among Spanish born in the second half of the 20th century. Likewise, the achievement of higher levels of education should be explained according to the main tenets of the human capital theory by the socioeconomic status of the student's origin family. The variable that better describes the socioeconomic status for the

generations studied is father's occupation, what will be also used as a valid instrument for the attainment of higher levels of education. These hypotheses will be tested and discussed in the results section.

3 Data Set and Variables

The survey where the data come from is the Living Conditions Survey, collected by the National Institute of Statistics in Spain. The Living Conditions Survey emerged as a continuation of the European Community Household Panel (ECHP). This survey contains information about households and single individuals. Although the total number of observations (N=29,211) is small in comparison with the population of the country, the sample is designed as a nation-wide representation of all the people.

Only wage-earners will be taken into account in this study because of several reasons. First, some researchers such as Vila and Mora (1998) believe that self-employed workers do not give reliable information about their own income. Second, if the net wage per worked hour is used as a measure of income to calculate the returns to education, it will be more appropriate to study wage-earners because self-employed workers choose their own time to work so it will be more difficult (or less precise) to calculate the wage per hour for this group. Third, as we can see in the fourth column of Table 3, wage-earners comprise 84 percent of the whole working age population so they represent a significant percentage in comparison with the other groups such as self-employed and others (16 and less than 0.01 percent respectively).

Table 3 shows the distribution of the labour force in Spain by gender. Out of all the population, there is 48 percent that are wage-earners (84 percent of employed population), 9 percent self-employed (16 percent of the employed population) and 0.02 percent other type of employees. We can see that the proportion of males participating in the labour force is larger than female participants (81 in comparison with 68 percent). In Spain the percentage of self-employed, of the employed population, for male and female (12 and 6 percent respectively) is very low in comparison with the high percentage of wage-earners, although superior to the northern and central european countries. Estimating the returns to education for the self-employed in Spain would not be very representative of the whole working population. To conclude the analysis of the labour force, the proportion of male wage-earners (51 percent) do not differ much to the proportion of wage-earners female (46 percent).

Table 4 contains the average net monthly earnings of wage-earners by gender and maximum education level attained. At all education levels, male earnings are significantly greater than female earnings. The gaps in earnings are not very different across education levels. The highest gap between males and females occurs in the secondary second stage with a difference in earnings of 424.3 euros in 2011. In relative terms, on average male earnings are 23 per cent higher than female earnings.

In many Spanish studies only male workers are considered to calculate the returns to education. This is to avoid the sample selection problems that female individuals can cause in the estimation. These problems arise from the fact that female careers are affected and sometimes stopped by caring duties or birth of children. Therefore, regarding the estimation, these differences could

cause sample selection issues. In this study female participants are also taken into account because there seem to be very similar numbers of males and females in this sample (out of all the wage-earners in the sample we have around 53 percent males and 47 percent females), so it is considered that there is no need to control for the bias that female participation can cause. It could be argue that it is necessary to control for male selection issues too.

The large number of variables are classified into five different categories: sociodemographic, education, labour, health, income, inter-generational transmission of poverty. Information is available from the year 2004 to 2013. Each year, the survey contains a special subsection with new variables added in each subsection. The year 2011 is used because the subsection called "intergenerational transmission of poverty" provides us with variables about the socioeconomic background of each individual. This special subsection focuses on individuals that are between 25 and 59 years old. The questions asked refer to the period when the individual was a teenager (14 years old). When family background variables are included in the regressions the analysis is restricted to individuals aged 25 - 59 years while if non of these variables are used, the estimation sample consists of individuals older than sixteen years and younger than 65 (age of retirement).

In order to derive a continuous variable for education, we follow the common approach:

$$\begin{aligned}\text{YEARS OF EDUCATION} = &\text{Year that highest level of education was attained}\\ &- \text{Year of birth} - 6\end{aligned}$$

Children in Spain start primary school at the age of 6. If after creating this variable following the formula above we look at the summary statistics we find that the minimum value is 0 years and maximum 52. Obviously this procedure is not taking into account if the individuals have stopped their education and years after go back, or if they resit some years for example. In order to avoid these discontinuities, a procedure is followed. In this database we also have a variable that reports the highest level of education attained that takes values: 0 no education at all, 1 if primary education, 2 secondary first stage, 3 secondary second stage and 4 higher education. The average[1] years of education for those who have less than 22 years of schooling (6 years of primary education, 4 secondary education, 2 years of pre-university education, 4 years of degree in university, 2 of Masters and 3 of PhD) is calculated. These values are tabulated with levels in the table below. Now, putting together all the individuals (the ones who have less than 22 years of education (21,20, etc.) and the ones with more than or equal to 22 years of education) we truncate the cases where the individuals have more than 21 years and we replace the number of years with the one from the table below according to the education level that they report. So, for example, for an individual that has 25 years of education and reported a secondary second stage level we assign them the value 12.33 rather than 25. The continuous variable for education is called Years of Education.

The variables used for the levels of education attained are:

- No education: the individual has not attained any level of education at all.

[1]The median years of education has been tried but does not alter the estimation results.

Level of Education attained	Mean of Years of Schooling
No Education	0
Primary	7.62
Secondary First Stage	9.50
Secondary Second Stage	12.33
Higher Education	16.41

- Primary: the individual's highest education attainment is primary education.
- Secondary First Stage: the individual's highest educational attainment is secondary first stage education or equivalent.
- Secondary Second Stage: the individual's highest educational attainment is secondary second stage education or equivalent.
- Higher Education: the individual's highest educational attainment is an undergraduate degree, postgraduate, etc. or equivalent.

In order to make the analysis simpler, the variables that represent the region where the individual is living in 2011 are grouped into categories:

- North West: Galicia, Principado de Asturias and Cantabria.
- North :País Vasco, Comunidad Foral de Navarra, La Rioja and Aragón.
- East: Catalua, Comunidad Valenciana and Baleares' Islands.
- South :Andalucía, Región de Murcia, Extremadura, Ciudad Autónoma de Ceuta, Ciudad Autónoma de Melilla and Canarias.
- Inner: Castilla-La Mancha and Castilla y León.
- Madrid: the capital of the country.

The following listed variables are used to estimate the wage equation:

- Experience: number of past years in paid work.
- Experience^2: experience squared (years squared).
- ln(wage): the logarithm of net wage per hour. The wage is measured this way because the distribution of log(hourly earnings) is close to a normal distribution. Heckman and Polachek (1974) tried several transformations of earnings and this one was the one finally chosen. Also it is convenient when it comes to interpreting the results.

And the last set of variables are the instruments for schooling (all these variables give information from the period when the respondent of the survey was a teenager):

- Father's Higher Education: dummy that takes the value of 1 if the father's highest level of education is higher education, 0 otherwise.

- Mother's education: The mother could have attained one of these levels:

  -No education: the dummy takes the value of 1 if the mother is illiterate, 0 otherwise.

  -Secondary First Stage: the dummy takes the value of 1 if the mother's highest educational attainment is secondary education, basic education or a level below this.

  -Secondary Second Stage: the dummy takes the value of 1 if the mother's highest educational attainment is post-secondary education.

  -Higher Education: the dummy takes the value of 1 if the mother's highest educational attainment is higher education.

- Father's Occupation: a recode of the OECD Occupational Classification ISCO-88 has been used grouping the occupation into three groups. The dummy that takes the value of 1 if the father had one of the following occupations, 0 otherwise.

  -Managers & Professionals: legislators, senior officials, managers, professional, technician, associate professionals.

  -Skilled Worker: clerk, service worker and shop and market sales worker, skilled agricultural and fishery workers, craft and related workers.

  -Unskilled worker: plant and machine operators and assemblers (elementary occupations).

- Public Expenditure. This is a macro economic series of the total public expenditure on education in Spain in the year when the individual was 12 years old. This variable is measured in millions of pesetas (currency used in Spain before the Euro). Data recorded on this variable from the year 1944 to 1994.

- Financial Difficulty. Economic difficulty for the family of making ends meet when the adult was a teenager. Takes the value of 1 if not difficult and 0 if difficult.

- Financial Situation. Economic household situation when the adult was a teenager. Takes the value of 1 if good financial situation and 0 if not good.

4 Methodology

Instrumenting Schooling to solve Endogeneity

In order to obtain consistent estimates of β_i in the wage equation when S and ε are correlated instrumental variables are needed in the model. The schooling equation is

$$S_i = \alpha_0 + \alpha_1 E_i + \alpha_2 E_i^2 + \alpha_3 IV_1 + \alpha_4 IV_2 + v_i$$

where: $E(v) = 0$, $Cov(E, v) = 0$, $Cov(E_i^2, v) = 0$ and $Cov(IV_k, v) = 0$ for $k = 1, 2$. To identify the model, the two coefficients, α_3 and α_4, have to be non zero.

Going back to the wage equation,

$$\ln(Y_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \varepsilon_i$$

we have the explanatory variables S, E and E_i^2 and the error term ε. It is assumed that: $E(\varepsilon) = 0$, the experience and its square are exogenous (not correlated with ε) but what is not sure is if S is also not correlated with the error term. It could be the case that such correlation exists and therefore $Cov(S, \varepsilon) \neq 0$.

Another test that is needed to be performed is the endogeneity test or Hausman test; To carry out this test this steps are followed:

1. The schooling equation (or reduced form) is estimated.
2. The residuals from the schooling equation are saved, $\hat{v}$;
3. The wage equation is estimated now but including these residuals as another regressor:

$$\ln(Y_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \varepsilon_i + \delta \hat{v}$$

4. The last step is to test the null hypothesis $H_0 : \delta = 0$. If the null hypothesis is rejected, the schooling variable is endogenous and therefore 2SLS method is to be used rather than OLS.

There are two possible scenarios: the schooling variable being exogenous, in which case OLS estimation will give consistent estimates and the schooling variable being endogenous, in which case OLS will give wrong estimates. In the second scenario, 2SLS is needed to overcome the endogeneity problem. This procedure comprises two steps:

1. In the first step, instrumental variables are used in order to predict the amount of schooling that each individual has accumulated. In order to explain the method, assume that two instrumental variables are used. The following equation is estimated:

$$S_i = \alpha_0 + \alpha_1 E_i + \alpha_2 E_i^2 + \alpha 3 IV_1 + \alpha_4 IV_2 + v_i \tag{2}$$

   Once equation 2 is estimated (called the reduced form), the predicted values $\hat{S}_i$ are saved.

2. In the second step, the wage $\ln(Y_i)$, is regressed on E_i, E_i^2 and $\hat{S}_i$ using OLS. If this procedure is followed, the coefficient on schooling would be a consistent estimate of the returns to education if the instruments are valid.

The most difficult task in the procedure is to find good instrumental variables. Each instrumental variable has to meet two criteria: in order to be valid instruments, they must not be correlated with the error term of the wage equation (ε) and they must be highly correlated with the endogenous variable, which in this case is schooling. In this study, the used instrumental variables are the following: the level of education of the parents, the public education expenditure in Spain and proxies of the economic situation of the household when the individual was a teenager. Depending on the form (discrete or continuous) of the endogenous variable different set of instruments will be used.

Testing the IV

In addition to the conventional tests for endogenenity and instrument validity another check is to estimate the model using Fuller-Limited Information Maximum Likelihood (Fuller- LIML) instead

of 2SLS. Fuller-LIML estimation is more robust than 2SLS. What is needed to be controlled is that the estimates on the returns of education once Fuller-LIML is used do not change too much in comparison with the ones obtained by 2SLS.

As Davidson and MacKinnon (1993) explain, Fuller proposed a modified LIML estimator where a is a positive constant. In comparison to the LIML estimator, that has no finite moments, Fuller's modified estimator has all the moments finite provided when the number of observations is large enough. Also the F test of the first stage (the schooling equation) must be checked. As Dickson (2009) says "the modified LIML estimator introduced by Fuller, with the Fuller parameter (a) set to one is regarded as most robust to any potential weakness of the instrument".

5 Results

5.1 Mincerian Earnings Equations

OLS estimations are reported in Table 5. Following Blackburn and Neumark (1993), industry and occupation controls are omitted for now because part of the returns to education can be captured by the variation in industry or occupation of employment.

As we can see, columns (a) and (c) are for the returns using the continuous measure of schooling, years, while columns (b) and (d) are for the returns using the discrete level of education attained. The coefficient on years of schooling gives us the estimated return to an additional year of education. The coefficients on the education levels gives us the return to the difference between not having education in comparison with higher levels of education. The drawback of using years rather than levels is that each additional year is assumed to have the same effect; one year of primary education will account for the same as one year of Masters studies. Comparing the rate of return among genders, males have lower returns to education than females. One more year of education will produce males wage increase in 5 and 6 percent for females. These results imply that there are gender difference in the returns to education in Spain. In the levels specification the reverse happens. Men have higher returns to secondary second stage education (42 percent for males and 17 percent for females) and higher education (88 percent for males and 71 percent for females). For females, the coefficients on levels primary and secondary first stage are not statistically significant. Having attained primary education or secondary education does not make a difference relative to females with no education at all. For males we observe that the coefficients on levels rise with the higher levels attained. Finally, there is more of an increase in the premium for females between the levels secondary second stage and higher education 38 percentage while for males there is an increase in the premium of 28 percentage.

5.2 Instrumenting Years of Schooling

As explained before, if the schooling variable is correlated with the error term, OLS estimates will be biased. Measurement error in the schooling variable may also bias the estimates downwards. Instrumental Variable estimation is a technique that controls for the possible bias due to omitted variables (such as ability) or measurement error.

Years of education are instrumented in Table 6 (Males) and Table 7 (Females). Three specifications for each gender are estimated: using fathers education , mothers education and both

instruments at the same time. Coefficients of the two stages of 2SLS are reported. The first stage is reported in columns (a), (c) and (e) while the earnings equation (second stage) is reported in columns (b), (d) and (f). When family background variables such as parents' education are used, only individuals aged 25-59 are taken into account (sample restrictions). Family background questions are only asked to individuals aged 25-59. Unfortunately, this subset of the population is not a random draw and this could lead to sample selection that cannot be controlled for.

All the instruments have the expected sign and are individually statistically significant. Parental education affects always positively own educational attainment. By observing the first stage coefficients we can say that the instruments are relevant.

The results suggest that the OLS coefficient on education is biased downwards. The change in the coefficients using OLS and 2SLS is for men 5.4 - 7.1 percentage points and for women 8.5 - 9.6 percentage points. We can see in Table 6, for males, fathers education has a larger impact in comparison with mothers education. This also happens for females if both instruments are included at the same time, column(c) of Table 7.

In order to see if the instruments are correlated with schooling a joint significance F test is performed for all the first stage. All the F tests statistics (Table 6 and Table 7) are above 10, so there is evidence that the instrumental variables used are relevant instruments and not potentially weak. When using both instruments, regarding the p-values, 0.895 and 0.353, from the overidentification tests, the null hypothesis of is not rejected. So we can confirm that when both instruments (mother's education and father's education) are used together the overidentifying restrictions are valid. The smallest partial R^2 of the effect of the instrument on years-of-schooling having partialled out the effect of the other covariates is 0.026 which is high relative to the guidelines given by Bound, Jaeger, and Baker (1995). The p-values associated with the endogeneity test are all less than 5 percentage points so it is suggested that is better to use 2SLS as estimation method rather than OLS. It is believed that education is endogenous rather than exogenous.

Wage equations are obtained using Fuller-LIML to check the robustness. For males (Table 8) and using father's education as an instrumental variable the coefficient on years of schooling has the same size using both estimation methods (Fuller-LIML and 2SLS). The same happens with females (Table 9). When using Fuller estimation method the coefficients on years of education also still remain statistically significant. As explained in the methodology section, estimating the regressions by using Fuller-LIML and checking that the coefficient on schooling do not vary is a robustness check.

Borrowing Murray (2006) words regarding instrumental variables: we need to avoid "bad, weak and ugly instruments". 'Bad' instruments are those correlated with the disturbances. 'Weak' instruments, are those little correlated with the troublesome explanator (in our case schooling). 'Ugly' instruments are those that yield results uninformative about what we are interested in.

It seems that parental education is a commonly used instrument for schooling for the Spanish case. Even if fathers education, mothers education or both instruments are used, the coefficient on the returns between the three specifications do not vary much (7.2 percentage points for males and 9.6 percentage points for females from Table 6 and Table 7).

Those individuals that have more ability are the ones who tend to acquire more education. Due to the non-observed ability, the ability is gathered by the error term in the regression. This will

cause the error term be correlated with the education variable. Basically, if more educated parents pass on to the children the tendency of acquiring more education but do not pass innate ability, parents education will likely be a valid instrument.

There are researchers that do not like parents' education as instrumental variables for education. They argue that parental education should be included as a control in the wage equation rather that using it as an instrumental variable. Barceinas Paredes et al. (2001) says that it could be argued that if a family is in a favourable economic condition, this will affect in a positive way the level of education of the children, but this also will affect in a direct way the future wage of the children so this will make the instruments invalid. Basically the idea is that parents with higher levels of education and, as a consequence, a higher level of income have a wide social network that enables the future work conditions of the children. Several papers still use family background variables as instruments for education, however.

Hoogerheide, Block, and Thurik (2012) use data from the 2004 German Socio-Economic Panel and Bayesian analysis to provide confidence in the use of family background variables as instruments in income regressions. They say that as however, if one's father's education affects these circumstances and attitudes, then it is implausible that one's own education would have no (or a smaller) effect. In other words, if education has a causal effect on earnings, then it is implausible that an additional year of education will benefit one's son or daughter, but not (or to a lesser extent) oneself.

5.3 Instrumenting Levels of Education

In this section the regression models are separated depending on levels of education attained. The instrumented education variable in this case is discrete.

First of all, three dummies will be created. The variable level of education can take one of these: Primary Education, Secondary First Stage, Secondary Second Stage and Higher Education. Each dummy will take a value of 1 if the individual has the higher level of education between two consecutive levels and then 0 if the individual has the level below. As can be seen, there are three dummies D_1, D_2 and D_3. The first one, D_1, will capture the individuals that have primary education and those who have attained secondary first stage; the second one, D_2, will capture the individuals that have secondary first stage and secondary second stage; the third one, D_3, will capture those who have attained secondary second stage and have attained higher education. A dummy with primary education versus no education is not created because there are not enough observations in the sample that have no education. In the whole sample of wage earners there is only 68 individuals that do not have education at all.

The criteria used to report the final set of instruments is:

- They have to be relevant (individually statistically significant) for the instrumented education level on its own.
- They have to be relevant for the instrumented education level when using both instruments.
- They have to pass the validity tests (F test and overidentification) and the endogeneity test.

The reason why the regressions for the instrumented levels of education are not divided by gender is because when years of education (a continuous variable) is used as the schooling vari-

able, the sample size is much larger (9131 observations from Table 1) than the sample used when estimating the model for education levels. For example, the model with the dummy for those individuals that have primary education versus secondary first stage would be estimated on 3106 observations. As mentioned before, when using any instrumental variable the sample is restricted to individuals older than twenty four and younger than sixty. The statistical significance of the coefficients depends much on the sample size used. If the regressions for the instrumented level of schooling are split, the models do not perform that well as when estimating a regression for both genders. That is the explanation why we only estimate one regression (including both genders) for each of the three dummies. Two regressions are ran for each dummy. OLS and IV will be used here. In the 2SLS approach, the endogenous variable will be the dummy for the level of education.

Table 1: Frequencies for the schooling variables (individuals aged between 25-59)

Schooling Variable	**Freq. (Observations)**
Years of Schooling	9131
No Education	68
Primary	971
Secondary First Stage	2135
Secondary Second Stage	2177
Higher Education	3811

In order to make the instrumental variable approach work, we need variables (instruments) that

- Explains part of the variation in education.
- Are not correlated with the unobservables (such as ability).

PRIMARY EDUCATION VS SECONDARY FIRST STAGE (D_1)

For the lower levels dummy, the set of instruments chosen are: the specific level of education of the mother when the individual was fourteen years old and the Spanish public expenditure on education in the year when the individual reached twelve.

Mother's education is a classic instrument for education and is used in this study to instrument this lower level. It is expected that Spanish mothers spent more time with her child than fathers did when the child is young. This could be explained because of the traditional family structure that applies for the generations analyzed: some mothers quit their jobs when they had a child or they stopped working when they got married so they devoted their time more to childcare duties while the father was the head of the household. Therefore is expected that the education of the mother would affect more the education of the offspring than father's education when the individuals in the sample were children. The coefficients on mother's levels of education have the expected sign. The coefficients increase in size: the higher the level of education achieved by the mother, the larger the effect on the education of the son/daughter (Table 10 column (a)). If the mother is more educated, it is more likely that the individual achieves secondary first stage level of education.

The reason why the age of twelve is chosen to get the data on public expenditure on education relies on the approximate age when the individuals are expected to finish primary education

and start secondary education. It would have been more accurate if the expenditure in education variable refers to a regional level rather than the national expenditure but unfortunately there is no old data of the expenditure in education at a regional level and the region where the individual lived when the individual was a teenager is also unknown. Bear in mind that the sample analysed comprises individuals subject to different education systems. Different education systems were introduced at a different year in each region. This means that the exact year when the individual goes into the next level of education and the region when it happened is unknown; therefore a raise of school living age is not a feasible instrument in this chapter.

In order to support the idea of choosing the national public expenditure on education when the individual was twelve as an instrumental variable a brief explanation is given. The national public expenditure in other years was tried without obtaining any statistically significant coefficient for the expenditure variable in the first stage equation. The variables tried were: the national public expenditure when the individual was born and the national public expenditure when the individual was six years old). None of them work well to instrument these lowest levels of education. The idea of using the public expenditure as a variable to explain education comes from Barceinas (2003). As he explains, when an economy is in recession, governments tend to diminish the national social expenditure. A part of the national social expenditure is devoted to education. Therefore the education budget is reduced. This adjustment could affect the individual decisions of continuing studying in the following way: a reduction of the national public expenditure could cause a decrease in the number of scholarships offered or a decrease in the amount of money invested in educational infrastructure and resources. According to this explanation, the expected sign for the coefficient should be positive. Table 10 column (c) shows that the public expenditure used on its own as an instrument for this level has a small size but statistically positive and significant coefficient.

Talking about robustness, as proved in columns (a) and (c) of Table 10 both instruments are separately valid. They are statistically significant in the first stage equation. For these low levels is better to use the mothers education disaggregated dummies indicating all the possible levels rather than using the dummy that takes the value of 1 if the mother has higher education and 0 otherwise. Using this dummy indicating if the mother has attained higher education to instrument these low levels does not work well because it is not statistically significant in the first stage schooling equation.

The size of the effect of the expenditure in education in the achievement of secondary first stage level when using both instruments is quite small but the sign of the coefficient is as expected. Mothers education is more directly related to the education of the child than the national public education expenditure. If the mother has higher education, this has a larger impact than if she has primary education again using both instruments. Again, using both instruments, if the government increases the education expenditure when the individual was twelve, this would have a positive effect in in the acquisition of secondary first stage education. Individuals who attain secondary first stage education level have a wage premium of around 39 percentage points in comparison with those that only get primary education.

If we have a look to the regression that includes both instruments, it can be seen that they all pass the tests for instrument validity (F test and overidentification). Out of several sets tried

to instrument these low consecutive levels, these are a combination of variables that are suitable together as instruments for this particular education levels. Further checks have being made for the validity of education expenditure as an instrument for the Spanish case. If 2SLS is used in the wage equation to obtain the coefficient for the education level, and we introduce the education expenditure in the wage equation as a control variable, the coefficient of education expenditure is not statistically significant. It can be seen that this variable do not affect directly wages (Table 10 column (b)). If we observe the standard error of the coefficient of education in column (f), when using both instruments, is lower (0.07) than when the single instruments are used, 0.12 in column (b) and 0.11 column (b) from Table 10. This means that the extra variation in education is more precise when both instruments are used.

To conclude the analysis of the low levels of education, several combinations of instruments were tried and do not pass the validity test when used together. Note that if fathers education is used instead of mothers education the coefficient on education in the wage equation is not statistically significant anymore (which contradicts Human Capital Theory). Instrument sets discarded include:[2] mother‘s education combined with father's occupation, type of household tenancy, number of children in the household younger than eighteen (equivalent to number of siblings), number of people in the household that worked or household composition. They do not serve as instruments to reflect the difference between attaining a primary education level and a secondary first stage because they are not significant for education in the first stage equation or they do not pass the validity tests.

SECONDARY FIRST STAGE VS SECONDARY SECOND STAGE (D_2)

For the middle level dummies, two specifications are presented. The first one, in Table 12, contains this set of instruments: mother's higher education and father's occupation. The alternative specification, in Table 14, contains the wage equations estimated by 2SLS using two single instruments that represent the financial status of the household when the individual was fourteen years old.

The difference between these two levels, in terms of the years that it takes to attain the higher level (secondary second stage education) is not that large. People who only have secondary first stage education have attended all the courses of compulsory education (no matter what education system applies to them). They are supposed to acquire basic cognitive skills and general knowledge. Instead, secondary stage holders go further with their education and they do the pre-university courses which take approximately two years.

The first specification chosen explains the difference between these two consecutive levels through the mother's higher education and father's occupation variables. The reason why the specific mother's education level is not used as it is in Table 11 is because, if the different dummies are used for the specific mother's education level, the model does not pass the overindetification test. The only specific mother's education attained level that can be used as an instrument is higher education. Please note that if father's education is used as an instrument for these levels, it is statistically significant in the first stage. The reason why it is not reported is that if it is going

[2]Instrument sets discarded available upon request.

to be combined with father's occupation, it can be argued that father's occupation could depend on the level of education achieved by him. Mother's higher education and father's occupation also work as a good set of instruments for the next levels of education (higher education versus secondary second stage versus). Why these instruments only explain variation in the attainment of higher levels could be explained by thinking that they do not impact the offspring's education until he/she becomes older and has to decide between just obtaining secondary second stage education or tertiary education. If mother's occupation is used as an instrument instead of father's occupation the endogeneity test is not passed. This could be due to sample size given the traditional low level of labour participation of Spanish women. When father's occupation is used, the regressions are on around 3500 observations while if mother's occupation is used, the regressions are reduced to 1000 observations because of missing values contained in the mother's occupation variable. This could be due to the fact that, as said before, for the generations studied, there is a large percentage of mothers that are housewives, so they do not report any occupation.

As can be seen in Table 12, the coefficient on the father's occupation dummies have the expected sign. It can be seen that father's occupations that require more education levels are the ones that have a larger effect on the offspring's own education. Fathers who are managers, professionals and skilled workers would have a positive larger impact than those fathers who hold an elementary occupation on the probability of the child getting further education (specifically higher levels than primary education). Further test are carried out on the support of father's occupation. Table 13 shows that if father's occupation dummies are included as a control variables in the wage equation, none of their coefficients are statistically significant. It can be said that father's occupation has no further independent effect on wages.

The second specification specification in Table 14, explains the difference between individuals that attain secondary first stage education and secondary second stage through just the economic characteristics of the household. The instruments are financial difficulty and financial situation. It can be seen that they are both equivalent. Using one or the other gives a similar coefficient for education, a wage premium of around 104 percentage points in column (b) and around 122 percentage points in column (d) from Table 14.

Regarding the statistical analysis, both coefficients on the financial household variables are statistically significant in the first stage. These financial variables are valid as single instruments for all the education levels. This means that the difference between attaining one level or another could be explained by the financial situation of each household when the adult was a teenager. This makes complete sense because if an individual wants to continue studying, he/she (or their families) needs enough economic resources to buy stationery and books, to pay for transport to commute if needed or to pay the annual fees if the school or education center is not fully owned by the state (public schools) not to mention the opportunity cost of keeping on studying. The reason why they are chosen to be reported to explain these levels is to show that even if the difference between the consecutive levels is short in terms of the number of years that it would take to acquire the highest level (secondary second stage in this case), nevertheless the variables still serves to explain the variation. If financial variables that describe the economic situation of the household when the adult was a teenager are used as instruments, the models pass all instrumental variables validating tests. They are individually statistically significant at the first stage and their

coefficients indicate that, if the family do not have financial difficulties to meet the ends and the financial situation is good, this would positively affect the probability of attaining the secondary second stage level.

Bear in mind that the Human Capital Theory continues to hold in the case of Spain if levels of education are used as schooling variables. The higher the level of education attained is, the higher the wage premium that an individual gets. This can be checked if the education coefficients are compared between the lowest levels, Table 10 column (f) where the education coefficient is 39 percentage points when using both instruments and Table 12 column (f) with a premium of 92 percentage points for having attained the secondary second stage level in comparison with the ones that attain secondary first stage.

Taken together, the results just presented suggest that both instruments, mother's higher education and father's occupation are relevant variables for the highest level (secondary second stage) because they are individually statistically significant in the first stage. Also when used together, they are still relevant for the difference in education. Financial situation and financial difficulty are also valid instruments for these levels. To conclude the analysis of the middle levels of education, instrument sets discarded include: father's education, father's higher education and type of household tenancy, number of children in the household younger than eighteen, household composition. They do not pass the validity tests to reflect the difference between attaining a secondary first stage education level and a secondary second stage.

SECONDARY SECOND STAGE VS HIGHER EDUCATION (D_3)

For the highest levels dummies, the set of instruments chosen are: mother's higher education and father's occupation when the adult was fourteen. It is expected that a combination of two variables about the family socioeconomic status, mother's higher education dummy and father's occupation, it is sufficient to explain the variation between obtaining one of these levels.

Table 15 column (a) shows that the mother's higher education instrument has the expected sign in the first stage regression. The fact that the mother has attained higher education would affect positively the probability that the son/daughter attains higher education too. According to MECD (2012), a third of all the students that pass the university access exam have a father and/or mother that has attained higher education. It is reasonable to think that parents' education will affect the offspring's higher education achievement because in Spain, individuals tend to stay at home studying their university degrees and therefore living at the parents home reduce the cost of studying. MECD (2010) shows evidence of a very low residential mobility on account of higher education in Spain. The corresidence of the students with their parents until the end of the university period influences the adult's decisions on education.

The signs of the coefficient of the instruments related to father's occupation are as expected. Father's occupation could proxy for the socioeconomic status of the family or how wealthy the family is. If the father is a manager or professional, his wage will be higher than a father that hold a skilled worker occupation or an unskilled worker occupation. Managers and professionals will earn more money and therefore the family will be wealthier. Wealthy families could spend more money on education of the children. Therefore, if the individual has a father whose occupation is

manager or professional, this increases the probability of achieving higher education in comparison with those whose father has a skilled worker occupation or an elementary occupation. This can be seen in column (c) and (e) of Table 15. Father's occupation pass all the robustness checks. As can be seen in Table 15 column (c) and (e) father's occupation is individually statistically significant in the first stage. If both instruments, mother's higher education and father's occupation, are used together, column (f) Table 15, the F test of jointly significance and the overidentification test is passed. As done for the lower levels, if father's occupation is included as a control variable in the wage equation, Table 16, it is not statistically significant so does not affect wages.

Bear in mind that if father's education is used as an instrument combined to the type of household tenancy when the adult was fourteen, these set also works and pass all the robustness checks. There is several combinations of instruments that are valid to explain the attainment of higher education but, according to further tests the best combination of instruments is mother's higher education and father's occupation. Instruments that are not valid to explain the achievement of higher education are: number of children in the household younger than eighteen, number of people in the household who worked when the adult was fourteen or household composition. They are not valid because they do not pass some validity test or they are not individually significant in the first stage.

The returns to attain a higher education level are 136 percentage points in comparison with those who attain only secondary second stage education. The higher education holders include graduates from several subjects as doctors, architects, engineers, etc. and also individuals that have done a postgraduate program such as Masters.

To conclude the results section, it is worth to say that if levels of schooling are instrumented using sets of variables as it is done in Tables Table 10, Table 12 and Table 15, there is also a downwards bias in the OLS coefficients for education (8.6, 15 and 39.7 percentage points respectively for each level) reported on Table 17. For the lowest levels, comparing the coefficient of education of the 2SLS from column (f) Table 10 with the first column in Table 17 the magnitude of the bias is of 30 percentage points. For the medium levels, comparing the coefficient of education of the 2SLS from column (f) Table 12 with the second column in Table 17 the magnitude of the bias is even larger, 77 percentage points. For the higher levels, comparing the coefficient of education of the 2SLS from column (f) Table 15 with the second column in Table 17 the magnitude of the bias is 97 percentage points. The bias in the schooling coefficients increase as the levels attained increases. The difference in the coefficients obtained by OLS and 2SLS is statistically significant. The p-value of the endogeneity test is always 0.000 meaning that the null hypothesis can always be rejected for all the levels of education. The null hypothesis of education been an exogenous variable could be rejected, therefore 2SLS is a more adequate method in comparison with OLS.

6 Conclusion

Several models are presented in this paper: Mincerian Earnings Equations to estimate the returns to education and models instrumenting schooling (separated specifications depending if the schooling variable is discrete or continuous).

Regarding the Mincerian Earnings Equations, men are worse off than women. Their return

to an extra year of education is around 1 percentage point lower than for females. However, the reverse happens when the coefficient on returns to an extra level of education is calculated. In the latter regression, males are better off than females.

Returns to an extra year of education are calculated for individuals aged 25-59. 2SLS approch is used as the estimation method. Parental education is used as an instrumental variable. Mother's education has a larger effect on the education of the offspring in comparison with the father's education. Again, the OLS education coefficients are biased downwards suggesting that there is a need to control for the measurement error.

Returns to education in Spain have been estimated. It is believed that parents education is a good instrumental variable to solve the endogeneity problem that arises if OLS estimation method is used to obtain the estimate of the returns. Using the Spanish sample, OLS estimates are biased downwards. If years of schooling are used as the schooling variable, female returns to education are always higher.

The preferred instruments for the levels specifications are: mother's education and public expenditure for the lower levels and mother's higher education and father's occupation for the middle and highest levels. The reason why they are preferred is because they pass all the validity test and they are good proxies for the socioeconomic status of the family when the individual was a teenager. So, to summarize, the socioeconomic status is the main predictor of the decision of attaining further levels of education.

References

ABADIE, A. (1997): "Changes in Spanish labor income structure during the 1980s: A quantile regression approach," *Investigaciones Economicas*, 21(2), 253–272.

ALBA-RAMIREZ, A. (1993): "Mismatch in the Spanish labor market: overeducation?," *Journal of Human Resources*, pp. 259–278.

ANGRIST, J. D., AND A. B. KRUEGER (1990): "Does compulsory school attendance affect schooling and earnings?," Discussion paper, National Bureau of Economic Research.

——— (1992): "Estimating the payoff to schooling using the Vietnam-era draft lottery," Discussion paper, National Bureau of Economic Research.

ARRAZOLA, M., AND J. DE HEVIA (2008): "Three measures of returns to education: An illustration for the case of Spain," *Economics of Education Review*, 27(3), 266–275.

ASLAM, M. (2007): "Rates of Return to Education by Gender in Pakistan.," .

BARCEINAS, F. (2003): "Endogeneidad y rendimientos de la educación," *Estudios Económicos*, pp. 79–131.

BARCEINAS PAREDES, F., ET AL. (2001): "Capital humano y rendimientos de la educación en México," .

BECKER, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, by Gary S. Becker,...* London.

BLACKBURN, M. L., AND D. NEUMARK (1993): "Are OLS estimates of the return to schooling biased downward? Another look," Discussion paper, National Bureau of Economic Research.

BLAU, F. D., AND L. M. KAHN (2000): "Gender differences in pay," Discussion paper, National bureau of economic research.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American statistical association*, 90(430), 443–450.

BOWLES, S., H. GINTIS, ET AL. (1976): *Schooling in capitalist America*, vol. 57. New York: Basic Books.

BRUNELLO, G., AND R. MINIACI (1999): "The economic returns to schooling for Italian men. An evaluation based on instrumental variables," *Labour Economics*, 6(4), 509–519.

CARD, D. (1999): "The causal effect of education on earnings," *Handbook of labor economics*, 3, 1801–1863.

CARRASCO, R., J. F. JIMENO, AND A. C. ORTEGA (2008): "The impact of immigration on the wage structure: Spain 1995-2002," .

CORUGEDO, I., E. G. PÉREZ, AND J. M. PAGES (1992): "Educación y rentas. Una aplicación a la enseñanza media en España: Una nota," *Investigaciones Económicas*, 16(2), 299–304.

DAVIDSON, R., AND J. G. MACKINNON (1993): "Estimation and inference in econometrics," *OUP Catalogue*.

DE LA FUENTE, A., AND J. F. JIMENO (2005): "The private and fiscal returns to schooling and the effect of public policies on private incentives to invest in education: a general framework and some results for the EU," Discussion paper, CESifo working papers.

DEARDEN, L., S. MACHIN, AND H. REED (1997): "Intergenerational mobility in Britain," *The Economic Journal*, pp. 47–66.

DENISON, E. F. (1962): "Sources of economic growth in the United States and the alternatives before us," .

DICKSON, M. (2009): "The Causal Effect of Education on Wages Revisited," IZA Discussion Papers 4419, Institute for the Study of Labor (IZA).

DÍEZ, M. M. G., M. J. P. VILLADÓNIGA, AND J. P. RODRÍGUEZ (2011): "El rendimiento del capital humano de los trabajadores extranjeros y nacionales en España," *Revista de educación no 355. La gestión estratégica de la Educación Superior: retos y oportunidades*.

FERRADA, L. M., AND P. ZARZOSA (2010): "Diferencias regionales en la participación laboral femenina en Chile," *Cuadernos de economía*, 47(136), 249–272.

GARCÍA PRIETO, C., A. MARTÍN ROMÁN, AND C. PÉREZ DOMÍNGUEZ (2005): "Actual and Potential Returns to Schooling in Spain," *Education Economics*, 13(4), 387–407.

GREENE, W. H. (1997): "Econometric analysis, 1997," .

GRILICHES, Z. (1977): "Estimating the returns to schooling: Some econometric problems," *Econometrica: Journal of the Econometric Society*, pp. 1–22.

HARMON, C., H. OOSTERBEEK, AND I. WALKER (2003): "The returns to education: Microeconomics," *Journal of economic surveys*, 17(2), 115–156.

HECKMAN, J., AND S. POLACHEK (1974): "Empirical evidence on the functional form of the earnings-schooling relationship," *Journal of the American Statistical Association*, 69(346), 350–354.

HECKMAN, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, pp. 153–161.

HERNÁNDEZ, L., AND L. SERRANO (2013): "Efectos económicos de la educación en España: una aproximacón con datos," *PIAAC*, p. 66.

HIDALGO, M. A. (2010): "A demand-supply analysis of the spanish education wage premium," *Revista de Economía Aplicada*, 18(54), 57–78.

HOOGERHEIDE, L., J. H. BLOCK, AND R. THURIK (2012): "Family background variables as instruments for education in income regressions: A Bayesian analysis," *Economics of Education Review*, 31(5), 515–523.

HOOGERHEIDE, L., AND H. K. DIJK (2006): "A reconsideration of the Angrist-Krueger analysis on returns to education," Discussion paper, Econometric Institute Research Papers.

LASSIBILLE, G., AND L. N. GOMEZ (1998): "The evolution of returns to education in Spain 1980–1991," *Education Economics*, 6(1), 3–9.

LEVIN, J., AND E. J. PLUG (1999): "Instrumenting education and the returns to schooling in the Netherlands," *Labour Economics*, 6(4), 521–534.

MECD (2010): "Datos y Cifras del Sistema Universitario Espaol: Curso 2010-2011," *Catálago de publicaciones del Ministerio: educacion.es*.

——— (2012): "Datos y Cifras del Sistema Universitario Espaol: Curso 2011-2012," *Catálago de publicaciones del Ministerio: educacion.es*.

MINCER, J. (1974): "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.," .

MURRAY, M. P. (2006): "The bad, the weak, and the ugly: Avoiding the pitfalls of instrumental variables estimation," *Available at SSRN 843185*.

OLIVER, J., J.-L. RAYMOND, J.-L. ROIG, AND F. BARCEINAS (1999): "Returns to human capital in Spain: A survey of the evidence," *Returns to Human Capital in Europe, Helsinki, The Research Institute of the Finnish Economy*.

PASTOR, RAYMOND, R., AND L. SERRANO (2008): "Es rentable invertir en capital humano en Espaa? Anlisis Regional e Internacional," *Investigaciones de Economía de la Educación*, 3, 163–172.

PINO, F. J. B., ET AL. (2014): "Rendimiento privado de la educación y prima salarial en España. Un meta-análisis," *Investigaciones de Economía de la Educación volume 9*, 9, 923–938.

PONS, E., AND M. T. GONZALO (2002): "Returns to schooling in Spain: how reliable are instrumental variable estimates?," *Labour*, 16(4), 747–770.

PSACHAROPOULOS, G. (1994): "Returns to investment in education: A global update," *World development*, 22(9), 1325–1343.

PUHANI, P. (2000): "The Heckman correction for sample selection and its critique," *Journal of economic surveys*, 14(1), 53–68.

RAMOS, M. I. A., AND D. GARCÍA-CRESPO (2003): "Desajuste educativo y salarios en España," in *Economía de la educación AEDE XII: 18 y 19 septiembre*, pp. 187–201. Universidad Carlos III.

RAYMOND, J. L. (2011): *Es rentable educarse? : marco conceptual y principales experiencias en los contextos espaol, europeo y en pases emergentes*, Estudios de la Fundacin de las Cajas de Ahorros / Serie economa y sociedad ; 53. Madrid : FUNCAS, Enth. 4 Beitr.

REQUENA, M., AND F. BERNARDI (2005): "El sistema educativo," *Tres décadas de cambio social en España, Madrid: Alianza*.

RODRÍGUEZ, M. A. D., AND C. ALBERT (2004): "Salud, salarios y educación," *Hacienda pública española*, (169), 11–34.

SALAS, M. (2002): "Cuatro décadas en Economía de la Educación," *Revista de Educación*, (328).

SCHULTZ, T. W. (1961): "Investment in human capital," *The American economic review*, pp. 1–17.

SPENCE, M. (1973): "Job market signaling," *The quarterly journal of Economics*, pp. 355–374.

STAIGER, D. O., AND J. H. STOCK (1994): "Instrumental variables regression with weak instruments," .

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business & Economic Statistics*, 20(4).

THUROW, L. C. (1975): "Generating Inequality: Mechanisms Of Distribution In The US Economy Author: Lester C. Thurow, Publisher: Basic Books Pa," .

TROSTEL, P., I. WALKER, AND P. WOOLLEY (2002): "Estimates of the economic return to schooling for 28 countries," *Labour economics*, 9(1), 1–16.

UUSITALO, R., S. HONKAPOHJA, AND E. KOSKELA (1999): *Essays in economics of education*. Department of Economics, University of Helsinki.

VILA, L.-E., AND J.-G. MORA (1998): "Changing Returns to Education in Spain during the 1980s," *Economics of Education Review*, 17(2), 173–178.

WEISBROD, B. A. (1962): "Education and Investment in Human Capital," *The Journal of Political Economy*, 70(5), 106–123.

Table 2: Sample Summary Statistics. Wage-Earners (17-65)

Variable	Mean	St. Dev	Min	Max
Experience	18.303	10.658	0	55
Experience^2	448.578	467.690	0	3025
Married	0.591	0.492	0	1
Income rest household	20509.420	18731.500	-40000	181373.8
Dependent Children	0.502	0.500	0	1
Employee	0.538	0.498	1	0
ln(wage)	2.148	0.425	0.069	4.080
Male	0.544	0.498	0	1
Public Education Expenditure	443291.9	348285.2	3855	1118000
Financial Difficulty	0.607	0.488	0	1
Financial Situation	0.724	0.446	0	1
EDUCATION				
No Education	0.019	0.136	0	1
Primary	0.112	0.315	0	1
Secondary First Stage	0.228	0.420	0	1
Secondary Second Stage	0.246	0.431	0	1
Higher Education	0.402	0.490	0	1
Years of Schooling	12.687	4.095	0	21
MOTHER'S EDUCATION				
No education	0.051	0.219	0	1
Secondary First Stage	0.825	0.380	0	1
Secondary Second Stage	0.068	0.252	0	1
Higher education	0.055	0.228	0	1
FATHER'S HIGHER EDUCATION	0.108	0.310	0	1
FATHER'S OCCUPATION				
Managers & Professionals	0.256	0.436	0	1
Skilled Worker	0.503	0.499	0	1
Unskilled Worker	0.240	0.427	0	1
REGIONS				
North	0.106	0.308	0	1
North West	0.089	0.285	0	1
East	0.305	0.460	0	1
South	0.231	0.421	0	1
Inner	0.096	0.294	0	1
Madrid	0.172	0.378	0	1

Notes: Summary statistics are based on weighted values.

Table 3: Distribution of Labour Force in Spain by Gender (16 - 64).

Status		Male		Female		Total	
		N	%	N	%	N	%
Unemployed	(a)	2,892.2	18.23	2,538.1	16.24	5,430.5	17.24
Employed	(b) = (c) + (d) + (e)	10,068.1	63.45	8,202.8	52.50	18,271.5	58.02
Self-employed	(c)	1,962.0	12.37	976.5	6.25	2,938.6	9.33
Wage-earners	(d)	8,102.3	51.06	7,222.6	46.22	15,324.9	48.66
Others	(e)	3.7	0.02	3.7	0.02	7.4	0.02
Total Labour Force	(f) = (a) + (b)	12,960.3	81.68	10,740.9	68.74	23,702.0	75.26
Out of Labour Force	(g)	2,906.5	18.32	4,884.7	31.26	7,791.2	24.74
All	(h) = (f) + (g)	15,866.8	100	15,625.6	100	31,493.2	100

Table 4: Average Monthly Earnings of Wage-Earners in Euros, by Education Level and Gender

Education Level / Gender	Male	Female	t-test (M-F)	
NO EDUCATION	1066.4	667.9	-3.49	***
	(65.83)	(87.16)		
PRIMARY	1225.6	823.0	-14.89	***
	(17.76)	(18.44)		
SECONDARY 1	1295.3	894.0	-21.19	***
	(13.24)	(12.06)		
SECONDARY 2	1499.2	1074.9	-17.85	***
	(18.03)	(15.04)		
HIGHER ED.	1970.8	1622.9	-12.94	***
	(22.11)	(16.22)		
ALL	1565.3	1268.0	-20.03	***
	(10.68)	(10.14)		

Notes: Standard errors of the mean are below the coefficients.

Table 5: OLS Mincerian Equations, by Gender (16-65) and Schooling.

Variables	Males				Females			
	Years (a) Coeff.		Levels (b) Coeff.		Years (c) Coeff.		Levels (d) Coeff.	
YEARS OF ED.	0.049	***			0.058	***		
	(0.00)				(0.00)			
EXPERIENCE	0.034	***	0.031	***	0.026	***	0.023	***
	(0.00)		(0.00)		(0.00)		(0.00)	
EXPERIENCE^2	-4.73E-04	***	-4.12E-04	***	-3.0E-04	***	-2.4E-04	***
	(0.00)		(0.00)		0.00		(0.00)	
PRIMARY			0.107	**			-0.062	
			(0.05)				(0.07)	
SECONDARY 1			0.201	***			0.019	
			(0.05)				(0.07)	
SECONDARY 2			0.353	***			0.161	**
			(0.05)				(0.07)	
HIGHER ED.			0.632	***			0.540	***
			(0.05)				(0.07)	
NORTH WEST	-0.110	***	-0.124	***	-0.060	***	-0.076	***
	(0.02)		(0.02)		(0.02)		(0.02)	
NORTH	0.017		-0.001		0.077	***	0.042	**
	(0.02)		(0.02)		(0.02)		(0.02)	
EAST	-0.066	***	-0.076	***	0.003		-0.007	
	(0.02)		(0.02)		(0.02)		(0.02)	
SOUTH	-0.125	***	-0.125	***	-0.01		-0.028	
	(0.02)		(0.02)		(0.02)		(0.02)	
INNER	-0.063	***	-0.077	***	0.017		-0.002	
	(0.02)		(0.02)		(0.02)		(0.02)	
CONSTANT	1.229	***	1.508	***	1.031	***	1.548	***
	(0.03)		(0.05)		(0.03)		(0.07)	
R^2	0.33		0.35		0.34		0.39	
N	4926		4936		4477		4480	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. Dependent Variable: LN_WAGE. NOEDUCATION and MADRID are the reference categories for education splines and regions respectively

Table 6: Instrumental Variable Estimation. Wage-Earners (Males), using Years of Schooling.

Variables	Father's education				Mother's education				Parent Education			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
YEARS OF ED.			0.071	***			0.072	***			0.072	***
			(0.01)				(0.01)				(0.01)	
EXPERIENCE	-0.120	***	0.038	***	-0.116	***	0.038	***	-0.108	***	0.038	***
	(0.02)		(0.00)		(0.02)		(0.00)		(0.02)		(0.00)	
EXPERIENCE^2	0.000		5.00E-04	***	1.09E-04		5.00E-04	***	1.05E-04		5.00E-04	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.740	**	-0.092	***	-0.984	***	-0.099	***	-0.697	**	-0.097	***
	(0.24)		(0.02)		(0.24)		(0.02)		(0.24)		(0.02)	
NORTH	-0.405	*	0.023		-0.538	**	0.022		-0.378	*	0.019	
	(0.22)		(0.02)		(0.22)		(0.02)		(0.22)		(0.02)	
EAST	-1.082	***	-0.049	**	-1.275	***	-0.052	**	-1.050	***	-0.052	**
	(0.21)		(0.02)		(0.21)		(0.02)		(0.21)		(0.02)	
SOUTH	-1.396	***	-0.104	***	-1.602	***	-0.104	***	-1.360	***	-0.106	***
	(0.21)		(0.02)		(0.21)		(0.03)		(0.21)		(0.02)	
INNER	-0.869	***	-0.046	**	-1.137	***	-0.047	*	-0.840	***	-0.049	**
	(0.24)		(0.02)		(0.24)		(0.03)		(0.24)		(0.02)	
FATHER'S HIGHER ED.	3.348	***							2.955	***		
	(0.20)								(0.21)			
MOTHER'S HIGHER ED.					2.956	***			1.487	***		
					(0.27)				(0.29)			
CONSTANT	15.278	***	0.883	***	15.637	***	0.880	***	15.069	***	0.88	***
	(0.29)		(0.10)		(0.29)		(0.14)		(0.29)		(0.09)	
Overidentification p-value			(-)				(-)				0.895	
Endogeneity p-value			0.000				0.007				0.000	
F-Test First Stage			287.00				116.265				154.181	
Partial R^2 of the IV			0.0626				0.026				(-)	
N			4307				4359				4274	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. FATHER'S ED and MOTHER'S ED are the instrumental variables; they are dummies for higher education of the father and the mother respectively.

Table 7: Instrumental Variable Estimation. Wage-Earners (Females), using Years of Schooling.

Variables	Father's education				Mother's education				Parent Education			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
YEARS OF ED.			0.096	***			0.091	***			0.096	***
			(0.01)				(0.01)				(0.01)	
EXPERIENCE	-0.113	***	0.031	***	-0.099	***	0.030	***	-0.105	***	0.031	***
	(0.02)		(0.00)		(0.02)		(0.00)		(0.02)		(0.00)	
EXPERICENCE^2	0.001		0.000	***	0.001		0.000	***	0.001		0.000	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.855	***	-0.039		-0.792	**	-0.045	*	-0.840	***	-0.042	*
	(0.23)		(0.02)		(0.23)		(0.02)		(0.23)		(0.02)	
NORTH	-0.347		0.089	***	-0.429	**	0.087	***	-0.365	*	0.090	***
	(0.21)		(0.02)		(0.22)		(0.02)		(0.21)		(0.02)	
EAST	-0.923	***	0.045	**	-1.094	***	0.038		-0.974	***	0.046	**
	(0.20)		(0.02)		(0.21)		(0.02)		(0.20)		(0.02)	
SOUTH	-1.171	***	0.039	*	-1.286	***	0.027		-1.177	***	0.037	
	(0.21)		(0.02)		(0.21)		(0.02)		(0.21)		(0.02)	
INNER	-0.727	**	0.060	**	-0.837	***	0.054	**	-0.726	**	0.059	**
	(0.24)		(0.03)		(0.24)		(0.03)		(0.24)		(0.03)	
FATHER'S HIGHER ED.	2.954	***							2.493	***		
	(0.19)								(0.21)			
MOTHER'S HIGHER ED.					2.996	***			1.585	***		
					(0.27)				(0.29)			
CONSTANT	15.318	***	0.424	***	15.428	***	0.510	***	15.205	***	0.426	***
	0.26		-0.11		0.27		-0.14		0.27		0.10	
Overidentification p-value			(-)				(-)				0.353	
Endogeneity p-value			0.000				0.000				0.000	
F-Test First Stage			250.33				127.26				137.31	
Partial R of the IV			0.0595				0.031					
N			3963				3991				3928	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. FATHER'S ED and MOTHER'S ED are the instrumental variables.

Table 8: Fuller LIML estimator. Males

Variables	Father's education				Mother's education				Parent Education			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
YEARS OF ED			0.071	***			0.071	***			0.072	***
			(0.01)				(0.01)				(0.01)	
EXPERIENCE	-0.120	***	0.038	***	-0.116	***	0.038	***	-0.108	***	0.038	***
	(0.02)		(0.00)		(0.02)		(0.00)		(0.02)		(0.00)	
EXPERIENCE^2	0.000		0.000	***	0.000		0.000	***	0.000		0.000	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.740	***	-0.092	***	-0.984	***	-0.099	***	-0.697	***	-0.097	***
	(0.22)		(0.02)		(0.23)		(0.02)		(0.23)		(0.02)	
NORTH	-0.405	**	0.023		-0.538	***	0.022		-0.378	*	0.019	
	(0.20)		(0.02)		(0.21)		(0.02)		(0.20)		(0.02)	
EAST	-1.082	***	-0.049	**	-1.275	***	-0.052	**	-1.050	***	-0.052	**
	(0.20)		(0.02)		(0.21)		(0.02)		(0.20)		(0.02)	
SOUTH	-1.396	***	-0.104	***	-1.602	***	-0.104	***	-1.360	***	-0.106	***
	(0.21)		(0.02)		(0.21)		(0.03)		(0.21)		(0.02)	
INNER	-0.869	***	-0.047	**	-1.137	***	-0.047	*	-0.840	***	-0.049	**
	(0.23)		(0.02)		(0.23)		(0.03)		(0.23)		(0.02)	
FATHER'S HIGHER ED.	3.348	***							2.955	***		
	(0.17)								(0.19)			
MOTHER'S HIGHER ED.					2.956	***			1.487	***		
					(0.22)				(0.24)			
CONSTANT	15.277	***	0.884	***	15.637	***	0.883	***	15.069	***	0.879	***
	(0.27)		(0.10)		(0.28)		(0.16)		(0.28)		(0.10)	
F-Test First Stage			380.22				180.78				219.06	
N			4307				4359				4274	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. FATHER'S ED and MOTHER'S ED are the instrumental variables.

Table 9: Fuller LIML estimator. Females

Variables	Father's education				Mother's education				Parent Education			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
YEARS OF ED.			0.096	***			0.091	***			0.096	***
			(0.01)				(0.01)				(0.01)	
EXPERIENCE	-0.113	***	0.031	***	-0.099	***	0.030	***	-0.105	***	0.031	***
	(0.02)		(0.00)		(0.02)		(0.00)		(0.02)		(0.00)	
EXPERICENCE^2	0.001		0.000	***	0.001		0.000	***	0.001		0.000	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.855	***	-0.039		-0.792	***	-0.045	*	-0.840	***	-0.042	*
	(0.22)		(0.02)		(0.23)		(0.02)		(0.22)		(0.02)	
NORTH	-0.347	*	0.089	***	-0.429	**	0.087	***	-0.365	*	0.090	***
	(0.20)		(0.02)		(0.21)		(0.02)		(0.20)		(0.02)	
EAST	-0.923	***	0.044	**	-1.094	***	0.037		-0.974	***	0.046	**
	(0.20)		(0.02)		(0.21)		(0.02)		(0.20)		(0.02)	
SOUTH	-1.171	***	0.039		-1.286	***	0.027		-1.177	***	0.037	
	(0.21)		(0.02)		(0.22)		(0.03)		(0.21)		(0.02)	
INNER	-0.727	***	0.060	**	-0.837	***	0.054	**	-0.726	***	0.059	**
	(0.24)		(0.03)		(0.24)		(0.03)		(0.24)		(0.03)	
FATHER'S HIGHER ED.	2.954	***							2.493	***		
	(0.14)								(0.16)			
MOTHER'S HIGHER ED.					2.996	***			1.585	***		
					(0.17)				(0.20)			
CONS	15.319	***	0.426	***	15.428	***	0.514	***	15.205	***	0.427	***
	(0.27)		(0.11)		(0.27)		(0.15)		(0.27)		(0.11)	
F-Test First Stage			449.78				302.99				281.81	
N			3963				3991				3928	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. FATHER'S ED. and MOTHER'S ED are the instrumental variables.

Table 10: Primary Education VS Secondary First Stage

Variables	Mother's Levels of Education				Public Education Expenditure				Both Instruments			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
SECONDARY 1 (D_1)			0.418	***			0.254	**			0.330	***
			(0.12)				(0.11)				(0.07)	
EXPERIENCE	0.013	***	0.007	**	0.015	***	0.013	***	0.023	***	0.008	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
EXPERICENCE^2	0.000	***	0.000		0.000	***	0.000		-0.001	***	0.000	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.088	***	-0.063	**	-0.095	***	-0.065	**	-0.087	**	-0.070	**
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
NORTH	-0.105	***	0.113	***	-0.087	***	0.098	***	-0.105	***	0.104	***
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
EAST	-0.154	***	0.029		-0.150	***	0.006		-0.161	***	0.015	
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
SOUTH	-0.156	***	-0.041		-0.158	***	-0.062	**	-0.162	***	-0.056	**
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
INNER	-0.047		-0.012		-0.047		-0.021		-0.047		-0.017	
	(0.03)		(0.03)		(0.03)		(0.02)		(0.03)		(0.03)	
EDUCATION EXP.					2.280E-07	***			2.940E-07	***		
					(0.00)				(0.00)			
MOTHER'S ED.												
Secondary 1	0.201	***							0.183	***		
	(0.03)								(0.03)			
Secondary 2	0.306	***							0.265	***		
	(0.06)								(0.06)			
Higher Ed.	0.328	***							0.269	***		
	(0.10)								(0.10)			
CONSTANT	0.584	***	1.473	***	0.638	***	1.556	***	0.320	***	1.543	***
	(0.06)		(0.10)		(0.05)		(0.09)		(0.07)		(0.07)	
Overidentification p-value											0.125	
Endogeneity p-value											0.000	
F-Test First Stage											27.974	
N			2816				3240				2816	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: SECONDARY 1. No Education is the reference category for the mother's level of education.

Table 11: Instrumental Variable Estimation using MOTHER'S ED., including EDUCATION EXP. as an Explanatory Variable.

Variables	Mother's Levels of Education			
	(a) First Stage		(b) IV	
SECONDARY 1 (D_1)			0.428	***
			(0.13)	
EXPERIENCE	0.023	***	0.005	
	(0.00)		(0.00)	
EXPERICENCE^2	-0.001	***	0.000	
	(0.00)		(0.00)	
NORTH WEST	-0.087	**	-0.063	**
	(0.03)		(0.03)	
NORTH	-0.105	***	0.114	***
	(0.03)		(0.03)	
EAST	-0.161	***	0.031	
	(0.03)		(0.03)	
SOUTH	-0.162	***	-0.039	
	(0.03)		(0.03)	
INNER	-0.047		-0.012	
	(0.03)		(0.03)	
EDUCATION EXP.	2.94E-07	***	-4.970E-08	
	(0.00)		(0.00)	
MOTHER'S ED.				
Secondary 1	0.183	***		
	(0.03)			
Secondary 2	0.265	***		
	(0.06)			
Higher Ed.	0.269	***		
	(0.10)			
CONSTANT	0.320	***	1.513	***
	(0.07)		(0.08)	
N			2816	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: SECONDARY 1. No Education is the reference category for the mother's level of education.

Table 12: Secondary First Stage VS Secondary Second Stage.

Variables	Mother's Higher Education				Father's Occupation				Both Instruments			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
SECONDARY 2 (D_2)			0.542	***			0.644	***			0.653	***
			(0.16)				(0.10)				(0.10)	
EXPERIENCE	-0.003		0.021	***	-0.003		0.021	***	-0.003		0.022	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
EXPERICENCE^2	0.000		0.000	***	0.000		0.000	**	0.000		0.000	**
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.013		-0.102	***	0.026		-0.124	***	0.021		-0.123	***
	(0.03)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)	
NORTH	0.003		0.028		0.032		0.008		0.029		0.007	
	(0.03)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)	
EAST	-0.025		-0.057	**	0.002		-0.074	***	-0.001		-0.072	***
	(0.03)		(0.02)		(0.03)		(0.02)		(0.03)		(0.02)	
SOUTH	-0.110	***	-0.046		-0.082	***	-0.057	**	-0.085	***	-0.053	*
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
INNER	-0.085	**	-0.019		-0.058	*	-0.041		-0.060	*	-0.039	***
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
MOTHER'S HIGHER ED.	0.292	***							0.203	***		
	(0.05)								(0.05)			
FATHER'S OCCUPATION												
Skilled Worker					0.060	***			0.061	***		
					(0.02)				(0.02)			
Managers & Professionals					0.200	***			0.185	***		
					(0.02)				(0.02)			
CONSTANT	0.606	***	1.478	***	0.520	***	1.440	***	0.517	***	1.430	***
	(0.04)		(0.10)		(0.04)		(0.07)		(0.04)		(0.07)	
Overidentification p-value											0.674	
Endogeneity p-value											0.000	
F-Test First Stage											27.284	
N			3911				3584				3541	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: SECONDARY 2. Instrumental Variables: MOTHER'S HIGHER ED. and FATHER'S OCCUPATION. The reference category for the last instrument gathers unskilled workers.

Table 13: Instrumental Variable Estimation (D_2) using MOTHER'S HIGHER ED., including FATHER'S OCCUPATION as an Explanatory Variable.

Variables	Mother's Higher Education			
	(a)		(b)	
	First Stage		IV	
SECONDARY 2 (D_2)			0.648	**
			(0.26)	
EXPERIENCE	-0.003		0.022	***
	(0.00)		(0.00)	
EXPERIENCE^2	0.000		0.000	**
	(0.00)		(0.00)	
NORTH WEST	0.021		-0.122	***
	(0.03)		(0.03)	
NORTH	0.029		0.007	
	(0.03)		(0.03)	
EAST	-0.001		-0.071	***
	(0.03)		(0.02)	
SOUTH	-0.085	***	-0.054	
	(0.03)		(0.03)	
INNER	-0.060	*	-0.040	
	(0.03)		(0.03)	
MOTHER'S HIGHER ED.	0.203	***		
	(0.05)			
FATHER'S OCCUPATION				
Skilled Worker	0.061	***	-0.012	
	(0.02)		(0.02)	
Managers & Professionals	0.185	***	0.000	
	(0.02)		(0.05)	
CONSTANT	0.517		1.439	***
	(0.04)		(0.14)	
N			3541	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: SECONDARY 2. The reference category for the last instrument gathers unskilled workers.

Table 14: Secondary First Stage VS Secondary Second Stage. Alternative instrumental variables.

Variables	Financial Difficulty				Financial Situation			
	(a) First Stage		(b) IV		(c) First Stage		(d) IV	
SECONDARY 2 (D_2)			0.713	***			0.799	***
			(0.13)				(0.14)	
EXPERIENCE	-0.004		0.021	***	-0.004		0.022	***
	(0.00)		(0.00)		(0.00)		(0.00)	
EXPERICENCE^2	0.000		0.000	**	0.000		0.000	**
	(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.019		-0.091	***	-0.015		-0.090	***
	(0.03)		(0.03)		(0.03)		(0.03)	
NORTH	-0.016		0.036		-0.009		0.035	
	(0.03)		(0.03)		(0.03)		(0.03)	
EAST	-0.037		-0.047	*	-0.033		-0.045	*
	(0.03)		(0.03)		(0.03)		(0.03)	
SOUTH	-0.102	***	-0.020		-0.103		-0.011	
	(0.03)		(0.03)		(0.03)		(0.03)	
INNER	-0.100	***	0.007		-0.093		0.014	
	(0.03)		(0.03)		(0.03)		(0.03)	
FINANCIAL DIFF.	0.104	***						
	(0.02)							
FINANCIAL SIT.					0.112	***		
					(0.02)			
CONS	0.567	***	1.367	***	0.545	***	1.314	***
	(0.04)		(0.09)		(0.04)		(0.10)	
Overidentification p-value			(-)				(-)	
Endogeneity p-value			0.000				0.000	
F-Test First Stage			41.125				42.216	
N			3963				3969	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: SECONDARY 2. Instrumental Variables: FINANCIAL DIFF. and FINANCIAL SIT.

Table 15: Secondary Second Stage vs Higher Education

Variables	Mother's Higher Education				Father's Occupation				Both Instruments			
	(a)		(b)		(c)		(d)		(e)		(f)	
	First Stage		IV		First Stage		IV		First Stage		IV	
HIGHER ED. (D_3)			0.759	***			0.916	***			0.859	***
			(0.11)				(0.09)				(0.08)	
EXPERIENCE	-0.005	**	0.034	***	-0.007	***	0.035	***	-0.005	**	0.035	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
EXPERIENCE^2	0.000		0.000	***	0.000		0.000	***	0.000		0.000	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.049	**	-0.089	***	-0.049	*	-0.068	***	-0.047	*	-0.079	***
	(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)	
NORTH	0.018		-0.005		0.021		0.002		0.023		-0.006	
	(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)	
EAST	-0.050	**	-0.034	*	-0.053	**	-0.019		-0.049	**	-0.027	
	(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)	
SOUTH	-0.026		-0.050	**	-0.029		-0.039		-0.021		-0.049	**
	(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)	
INNER	-0.010		-0.030		0.001		-0.025		0.007		-0.033	
	(0.03)		(0.02)		(0.03)		(0.03)		(0.03)		(0.02)	
MOTHER'S HIGHER ED.	0.201	***							0.151	***		
	(0.02)								(0.02)			
FATHER'S OCCUPATION												
Skilled Worker					0.064	***			0.061	***		
					(0.02)				(0.02)			
Managers & Professionals					0.179	***			0.159	***		
					(0.02)				(0.02)			
CONSTANT	0.753	***	1.392	***	0.704	***	1.256	***	0.680	***	1.306	***
	0.03		(0.09)		(0.03)		(0.08)		(0.03)		(0.07)	
Overidentification p-value											0.261	
Endogeneity p-value											0.000	
F-Test First Stage											64.926	
N			5486				5148				5089	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: SECONDARY 2. Instrumental Variables: MOTHER'S HIGHER ED. and FATHER'S OCCUPATION. The reference category for the last instrument gathers unskilled workers

Table 16: Instrumental Variable Estimation (D_3) using MOTHER'S HIGHER ED., including FATHER'S OCCUPATION as an Explanatory Variable.

Variables	Mother's Higher Education			
	(a)		(b)	
	First Stage		IV	
HIGHER ED. (D_3)			0.697	***
			(0.15)	
EXPERIENCE	-0.006	**	0.034	***
	(0.00)		(0.00)	
EXPERIENCE^2	-0.000		-0.000	***
	(0.00)		(0.00)	
NORTH WEST	-0.048	*	-0.086	***
	(0.02)		(0.02)	
NORTH	0.023		-0.001	
	(0.02)		(0.02)	
EAST	-0.048	**	-0.033	
	(0.02)		(0.02)	
SOUTH	-0.021		-0.052	**
	(0.02)		(0.02)	
INNER	-0.007		-0.028	
	(0.03)		(0.02)	
MOTHER'S HIGHER ED.	0.150	***		
	(0.02)			
FATHER'S OCCUPATION				
Skilled Worker			-0.004	
			(0.01)	
Managers & Professionals			0.323	
			(0.03)	
CONSTANT	0.666	***	1.425	***
	(0.04)		(0.11)	
N			5089	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions. Variable instrumented: HIGHER ED. The reference category for the last instrument gathers unskilled workers.

Table 17: OLS with Schooling as consecutive education levels

Variables	D_1 Coeff.		D_2 Coeff.		D_3 Coeff.	
EDUCATION (D_i)	0.083	***	0.140	***	0.332	***
	(0.01)		(0.01)		(0.01)	
EXPERIENCE	0.011	***	0.020	***	0.031	***
	(0.00)		(0.00)		(0.00)	
EXPERIENCE^2	0.000		0.000	***	-0.000	***
	(0.00)		(0.00)		(0.00)	
NORTH WEST	-0.090	***	-0.119	***	-0.111	***
	(0.03)		(0.02)		(0.02)	
NORTH	0.081	***	0.014		0.002	
	(0.03)		(0.02)		(0.02)	
EAST	-0.025		-0.078	***	-0.056	***
	(0.02)		(0.02)		(0.02)	
SOUTH	-0.097	***	-0.105	***	-0.068	***
	(0.02)		(0.02)		(0.01)	
INNER	-0.030		-0.078	***	-0.043	**
	(0.03)		(0.02)		(0.02)	
CONS	1.738	***	1.746	***	1.729	***
	(0.04)		(0.03)		(0.02)	
R^2	0.117		0.173		0.273	
N	2816		3541		5089	

Notes: *, **, *** means p-value <0.10, p-value <0.05 and p-value <0.010 respectively. MADRID is the reference category for the regions.