Return on education around the European Union: A reappraisal

Domenico Depalo *

January 4, 2017

Abstract

The primary aim of this paper is to show that standard estimators that have been used to compare outcomes across countries are inadequate. As a leading example, I estimate the return on education in the EU countries. The OLS recovers the return on education for the population on average, but is not robust to possible endogeneity of schooling achievement. The IV is robust to endogeneity, but recovers the return on education only for compliers, that are different across countries and cannot be identified from observable data. Bounds are robust to endogeneity of schooling (unlike the OLS) and is valid for the population as identified by the data (unlike the IV). Return on education are quite different across countries. A useful interpretation of results is in terms of demand/supply schedule. A closer inspection of bounds allows to go a step further in explaining the differences in results.

*I thank Monica Andini, Caroline M. Hoxby, Marco Leonardi, Jeffrey Wooldridge. I also thank seminar participants at the Bank of Italy. All the routines will be available at the webpage: http://sites.google.com/site/domdepalo/ The views expressed in this paper do not imply any responsibility of the Bank of Italy. Author: Domenico Depalo, Banca d'Italia, Economics and Statistics Department, Via Nazionale, 91 - 00184 Rome (Italy), Tel.: +39-06-4792 5989, e-mail: domenico.depalo@bancaditalia.it

1 Introduction

The primary aim of this paper is to show that classical estimators that have been used to compare outcomes in different countries are not adequate, because in general they do not identify the parameter of interest. As a leading example to show the relevance of this paper, I estimate the return on education across countries belonging to the European Union (EU). It should be clear at the outset that the technical issues arisen in this paper are of general interest for the empirical research.

Wage return on education is a key policy indicator within a single country and across different countries. The investment in schooling is (economically) more convenient if returns are higher. To the extent that education increases individual and aggregate productivity, it fosters economic growth of the countries (Sianesi and Van Reenen (2003)). In this respect, return on education tells much of the economic strengths and weaknesses of a country.

Although the wage return on education has been largely investigated for several countries, there is little consensus not only about its size, but also about its statistical significance, i.e. its existence. For Germany, Trostel et al. (2002) estimate a positive return of 4–7% per year of education, whereas Pischke and von Wachter (2008) and Kamhöfer and Schmitz (2015) further qualify that the return is non significant for compulsory education. For higher education versus anything less in UK, the estimated wage return each additional educational year is 11-15% in Trostel et al. (2002) compared to 27% in Blundell et al. (2005). Not only do these differences reflect different reference populations and/or reference periods, but also different identification assumptions. Whilst the former differences can be easily explained, the latter is more tricky.

This paper critically reviews the role of maintained assumptions that are necessary to point identify the parameters of interest. Even though estimators are increasingly more refined to account for more realistic aspects of the wage determination, ultimately the estimated parameters depend on imposed hypotheses, that often are untestable (Card (1999), Blundell et al. (2005), Harmon et al. (2003) for an extensive discussions). Then the credibility of estimated parameters becomes an issue when confronted to the 'law of decreasing credibility' (Manski (2011)) that the credibility of inference decreases with the strength of the assumptions maintained.

If the schooling attainment is as good as randomly assigned, the OLS recovers the parameters of interest for the whole population. If schooling achievement depends on unobservable ability or heterogeneous private return, the OLS is inconsistent and other estimators must be employed. Typical solutions involve an IV estimator, imposing exogenous variation of the instrument. Under suitable conditions, IV identifies the treatment effect only for the subpopulation of compliers, i.e. a Local Average Treatment Effect (LATE). The compliers cannot be identified from observed data and their treatment effect can not be generalized to the entire population, unless under special circumstances. Therefore, using IV does not provide a meaningful parameter for a comparison across countries, because it targets different populations.

To solve these drawbacks, in this paper I take a different perspective with respect to the standard literature, by exploiting only the assumptions that may be tested and that are indeed verified in the data. Almost always, this implies abandoning point identification in favour of set identification (or bounds; Manski (1990) and the following literature). For an international comparison, set identification brings about two notable advantages over the classical approach: differently from the OLS, it is robust to possible non-random schooling attainment; differently from IV, it is valid for the entire population identified by the observational data rather than for a specific, unknown, subpopulation.

Return on education are quite different across EU countries. A useful interpretation of results is in terms of demand/supply schedule. The small middle-low return is consistent with the intermediate educational level being the highest share of the working age population, and thus the least remunerated. The larger high-middle return reflects the relative scarcity of highly educated individuals. The only exception to this argument is Italy where both the share of highly educated individual and return on highest education are among the lowest across countries analyzed in this paper (Visco (2014, 2015) has defined this as a 'paradox'). I leave for further research a better understanding of this paradox.

This paper is organized as follows. In Section 2, I describe the empirical strategy that is adopted to EU-SILC data (Section 3). The returns on education are discussed in Section 4 and some robustness checks in Section 5. Section 7 offers some concluding remarks.

2 Empirical strategy

The aim of this section is to make clear that standard estimators are not adequate when the interest is in comparisons across countries. Using set identification jointly solves the open issues.

As a leading example, I focus on the identification and estimation of the return to education in countries belonging to the EU. The exercise is difficult, because not all necessary information is available to the researcher. Various approaches have been proposed in the literature to solve the problem. In this section, I consider only key aspects that are relevant for an international comparison. The reader interested in further technical details should refer to Card (1999); Blundell et al. (2005); Blundell and Dias (2009) and the references therein. To make the exposition simpler, but without loss of generality, in this section I do not explicitly condition on observable characteristics. However, the analysis that follows should be understood as conditional on them.

Following Card (1999), assume that individuals maximize the utility function $U(S, y) = log(y) - h(S)$, where y is the individual earning, S the educational level, and $h(\cdot)$ an increasing convex function. This economic background is coherent with the Mincer (1974) equation $y_i = \beta S_i + u_i$, where $i = 1, \ldots, N$ represents the individuals, β is the return on education and u is the error term. In general, the optimal level of schooling is heterogeneous across individuals due to heterogeneity in benefits –represented by the marginal return on education– and costs. For this reason, the Mincer equation should be augmented to account for individual specific intercepts of the earning equation and/or heterogeneous slope of the earning-schooling relation. More formally, as in Blundell et al. (2005), define $u_i = \alpha_i + (\beta - b_i)\, S_i + \epsilon_i$, where α_i captures unobservable individual characteristics, like ability, $(\beta - b_i)$ is the individual-specific slope of the extra-return on education and ϵ is a random error term. The complete reference model thus may be written as

$$y_i = \beta\, S_i + \alpha_i + (\beta - b_i)\, S_i + \epsilon_i. \tag{1}$$

Different assumptions on components of equation 1 drive the choice of the appropriate estimator.

The simplest possible approach imposes the orthogonality condition that $S_i' \alpha_i = S_i'(\beta - b_i) S_i = 0$, so that the average return for the population who reached a certain level of education can be recovered by comparing the average earning of individuals who reached that level to those individuals who didn't reach it, as the OLS does. To better reflect that this is the effect for the population on average, from now on I label the estimator 'OLS-ATE'.

The utility maximization and eq. 1 suggest that the optimal educational level may be related to unobservable (to the researcher) characteristics, in which case the individuals reaching the educational level S will not be a random sample (i.e., the selection will no longer be ignorable; Little (1995); Little and Rubin (1986)) and the OLS-ATE will be biased. In particular, by inspection of the orthogonality conditions, the OLS-ATE may be biased due to: 1) non-zero correlation between α_i and schooling, e.g. if higher ability workers reach higher educational level, which induces an upward bias; 2) non-zero correlation between the private return $(\beta - b_i)$ and schooling, which implies a bias in an unpredictable direction.

Solving these drawbacks has been central in the literature (Card (1999)). The classical solution is based on instrumental variable (IV) techniques. The identification of β with IV hinges on a suitable instrument (Z) that must be correlated with education, i.e. relevant, but not with ability, i.e. exogenous $(\alpha' Z = 0)$.[1] For the moment, suppose there exists a Z that satisfies these conditions. Even in this case, IV will not recover the effect for the entire population, because $E[y_i|Z_i, X_i] = \beta E[S_i|Z_i, X_i] + E[(\beta - b_i) S_i|Z_i, X_i]$. The last term is different from zero, unless either return is homogeneous or, focusing on treated individuals, the instrument is uncorrelated also with the individual extra-return on education (i.e., the schooling decisions is unrelated to the individuals-specific return). Whether one of the two assumptions can be invoked should be evaluated case-by-case. However, it is worth emphasizing that both are particular cases, that may be unrealistic in the real world (Heckman (1997)).

A solution to treat the heterogeneous return is to augment the standard exclusion restriction

[1] In this section I do not consider the control function (CF) approach (see Wooldridge (2015)). Both IV and CF are asymptotically consistent if all hypotheses are met, in which case CF has better properties than IV (e.g., in terms of efficiency). However, IV is consistent under weaker conditions, thus is preferred here instead of CF. The weaker requirement for consistency of IV is also related to the different identification power of the two estimators, an issue that is tackled below. Blundell and Dias (2009) provide an extensive discussion of other difficulties with CF.

and relevance assumptions with monotonicity. Monotonicity implies that there is no one who does the opposite of his/her assignment of instrument. Under these hypotheses, IV identifies a *Local* Average Treatment Effect. To better reflect the local identification power, from now on I label the estimator 'IV-LATE'. It is the effect for the *sub*-population of compliers, individuals who are induced to invest in education by a change in the instrument (Imbens and Angrist (1994)): different instruments affect different subpopulations of compliers each -in general- with different returns. The understanding of which sub-population of compliers is induced to take more schooling should be driven by an economic framework. To check the framework, one would like to individually flag the compliers. This is not possible because they are not identified from observed data (Angrist et al. (1996)), although the distribution of their characteristics can be described (or 'characterization of compliers'; Angrist (2004) and Angrist and Pischke (2008)).

The identification of a *local* effect rather than for the population at large and the impossibility to flag the compliers is a major limit of IV-LATE when the interest is in comparisons across countries. $Different$ instruments target different populations, both within and between countries, therefore being inappropriate for the scope. More subtle -and never noticed in the existing literature- is that in general even the *same* instrument identifies sub-populations of compliers that are different from country to country. The reasons may be due, for example, to different cultural background or institutions. To be more concrete, consider using the adverse financial shock experienced by the family when the child was 11 or 16 years old, used by Blundell et al. (2005) to estimate the return on education in UK. If the (perceived) quality of the public versus private school is different across countries, this instrument will target a different population of compliers: in countries where the quality of public school is higher than the quality of private school and where the age limit for compulsory school is 16 or above, the instrument will be less relevant than in countries where the quality of private school is higher than the quality of public school or where the age limit is lower. With these differences, finding an instrument that accommodates all the countries is hard *if ever possible.*[2] This property implies that, while being robust to possible endogeneity of schooling

[2] By the same token, the strategy proposed by Ichino and Winter-Ebmer (1999) to bracket the return on education *in the population* of Germany, using two different instruments that target the highest return on education (i.e., the most able workers with binding balance constraints) and the lowest return on education (i.e., the least able with the weakest balance constraints) does not work for an international comparison.

attainment, a fair comparison across countries based on IV-LATE is no longer correct, unless when the parameter is identified for the entire population. This happens in very special cases when the return is homogeneous or when self-selection does not depend on expected heterogeneous return, as pointed out in Blundell et al. (2005) for the return on education and in Heckman (1997) for a more general case.

With IV-LATE there is one more difficulty. So far I have assumed I have a valid instrument. However, while standard instruments to estimate the return on education involve family background (usually parental education), there is no guarantee that it satisfies all necessary conditions for consistent estimates. To the extent that family background is related to wage, maybe only indirectly through ability, the IV-LATE estimate will be upward biased, more biased than the OLS-ATE. Card (1999) shows that in this case $\beta \leq \beta_{OLS} \leq \beta_{IV}$, where β is the true parameter and the subscript is for the estimator.

To circumvent critical issues of classical estimators, related to *i)* the role of assumptions and *ii)* the characterization of the population, it may be useful to investigate other approaches. In this paper, I exploit bounds (Manski (1990)). They identify a set (and not a single point) of admissible marginal effects that depend on the underlying assumptions. Therefore, rather than imposing untestable assumptions, I retain only those that are verifiable and that are met by the data. Related to the present analysis, Manski and Pepper (2000, 2009) emphasize that instruments' validity is often a matter of disagreement, therefore it is worth considering weaker, but more credible, assumptions.

The rationale beyond bounds is that each member of the population with observable characteristics $x_i \in X$ is exposed to a mutually exclusive and exhaustive treatment $s \in S$ such that $y_i(s) \in Y$. The outcome under realized treatment is observable. The latent outcome is not observable. Combining the empirical evidence with assumptions, one may learn about the distribution of the response function. For example, consider the wage level after s_i year of education. Then $E[y(s)] = E[y|t = s]\,P(t = s) + E[y|t \neq s]\,P(t \neq s)$. The quantity $E[y|t \neq s]$ is not observed, but if I know that $E[y] \in [K_0, K_1]$ then $E[y|t = s]\,P(t = s) + K_0\,(1 - P(t = s)) \leq E[y(s)] \leq E[y|t = s]\,P(t = s) + K_1\,(1 - P(t = s))$. The larger the set of assumptions on the unobserved components, the narrower the bounds.

Continue with the estimation of return on education. If individuals of higher ability enjoy a higher mean wage function than those with lower ability, the so called Monotone Treatment Selection (MTS) is a promising alternative to the standard IV-LATE assumptions. More formally, whilst the IV-LATE requires that $E[y(s)|Z = u_2] = E[y(s)|Z = u_1] \, \forall s \in S$, MTS implies $E[y(s)|Z = u_2] \geq E[y(s)|Z = u_1]$. The human capital theory, suggests also that wage is a weakly increasing function of the years of schooling, or Monotone Treatment Response (MTR). Formally, MTR assumes that for $\{s_1, s_2\} \in S$, $s_2 \geq s_1 \Rightarrow y_i(s_2) \geq y_i(s_1)$. MTR and MTS are not mutually exclusive, and can be exploited together as I do in Section 4.3, in which case

$$\sum_{u<s} E[y|z = u] \, P(z = u) + E[y|z = s] \, P(z \geq s) \leq E[y(s)] \leq \sum_{u>s} E[y|z = u] \, P(z = u) + E[y|z = s] \, P(z \leq s).$$

Having defined bounds for $E[y(s)]$, also the education return $\Delta(s_1, s_2)$ can be bounded: its upper bound is the difference between the upper bound of s_2 and lower bound of s_1. In contrast, lower bound is set to zero (i.e., more schooling is not harmful on individuals' earning). A great advantage of this estimation is in terms of the law of decreasing credibility, because the two hypotheses can be jointly tested. Namely, the hypotheses should be rejected if $E[y|t = s]$ is not weakly increasing in s (Note 9 in Manski and Pepper (2000)).

Having described all the estimators, it is possible to compare them with respect to a comparison across countries. Provided the working hypotheses are verified, bounds are robust to possible endogeneity (e.g., non random schooling achievement), unlike the OLS-ATE. Also, with respect to characterization of the population, bounds apply to the population as defined by the conditioning set (Manski and Pepper (2000)) and not to subpopulations of unknown composition, as defined by local estimators like IV-LATE. In all the economic problems where the interest is in comparison across countries, using bounds guarantees the comparison of consistent parameters of interest along the very same populations, across different countries. This implies that any difference is due only to the national economic system and not to other confounding effects. Table 1 summarizes the properties of all these estimators.

The literature on return on education is huge. Among the first international comparison is Psacharopoulos (1985). Card (1999); Harmon et al. (2003) are very useful surveys for methods and results on return on education around the world. The technical points of this paper go well beyond retunr on education. Still on human capital, recent international investigations involving schooling and ability are in Hanushek and Woessmann (2011); Hanushek et al. (2015).

A review of bounds is in Ho and Rosen (2015). related to endogeneity and schooling, Manski and Pepper (2000, 2009) to estimate the return on education in the USA. Bounds closely related to return on education are Ginther (2000) and Mariotti and Meinecke (2015) that focus on a single country. To the best of my knowledge, this is the first paper that exploits the precise characterization of the population for cross-countries comparison of the return of education.

3 Data and descriptive statistics

The practical relevance of Section 2 is shown by estimating return on education in 25 European Union countries during the period 2004-2012.[3] Data are taken from the survey 'European Union Statistics on Income and Living Conditions' (EU-SILC), which collects timely and comparable data on a large array of socio-economic indicators. The sample is nationally representative of the population residing in private households aged 16 and over, irrespective of language, nationality or legal residence status. I exclude self-employed and women to limit possible bias arising from self-selection in the labour market participation. The reference population is made of men in the age range 25–65, which I conventionally define as the working age population.[4]

There are three relevant issues to be considered: 1) the schooling attainment, 2) the outcome of interest and 3) the possible instrument(s).

The educational level is based on the the ISCED level attained, which distinguish among pre-primary education, primary education, lower secondary education, (upper) secondary education, post-secondary non tertiary education and tertiary education.[5] In this analysis, the educational

[3] The countries are: Austria, Belgium, Cyprus, Czech Republic, Germany, Denmark, Estonia, Spain, France, Greece, Hungary, Ireland, Island, Italy, Lithuania, Luxembourg, Latvia, The Netherlands, Norway, Poland, Portugal, Sweden, Slovenia, Slovak republic and United Kingdom.

[4] Women are included in a robustness check in Section 5.

[5] Official documentation on the International Standard Classification of Education (ISCED) is available on the EU

attainment of a person is the highest level of an educational program the person has successfully completed; persons who have not completed their studies are coded according to the highest level they have completed. The expression 'level successfully completed' must be associated with obtaining a certificate. Therefore, results are valid only for official education. Having a reference in terms of grades rather than in terms of age or years of schooling provides several advances in this exercise. Enrollment ages and minimum durations are different from country to country (European Commission (2015)), thus with the alternative definitions would be difficult to disentangle the 'contribution of legal rules' from the true return on education. Also, return on education based on grades rather than on years of education is not affected by the so called 'sheepskin effect' (that instead should be modeled with the latter indicator) and might be more robust to measurement errors based on recall.[6] The middle educational attainment is the highest share in almost all the countries: exceptions are Spain and Portugal (where the lowest educational attainment has a larger share; in Spain the share of middle highest educational level is the smallest) and Belgium and Ireland (where the highest educational attainment has a larger share). Apart from Italy, Luxembourg and Portugal, the second highest share is for the highest educational level. Finally, it is worth emphasizing that the share of the highest educational attainment is usually higher than 20%: the only expcetions are Hungary, Italy, Poland, Slovenia and Portugal.

The outcome of interest is the (natural logarithm of) the hourly wage. I consider "gross monthly earnings for employees", which refers to the monthly amount of money received by the employee in his main job. For Germany, France and Slovenia, for which this variable is not available, I use employee "cash or near cash income": in this case, the variable is the sum of earnings from all jobs in the reference period, thus I restrict the analysis to individuals who have only one job, in order to avoid spurious relations. The hourly wage is calculated by dividing the employees' gross monthly earnings by the hours they usually work each week (multiplied by 4). The hourly wage for "employee cash or near cash income" is calculated accordingly. For each country, Table 2 reports the average hourly wage levels, by educational level, along with the proportion of sample found in

website.

[6] Of course, if one is specifically interested in the sheepskin effect, then this indicator is not appropriate. As for measurement error, it will be further considered in Section 4.2.

that level. The wage is increasing in education in all the countries, although with huge cross-country heterogeneity: the difference between wage at middle educational level and at low educational level (for short 'the middle-low wage difference') goes from 8% in Belgium to 35% in Luxembourg, with an average of 18% pooling all the countries; the high-middle wage difference ranges from 15% in Sweden to 65% in Hungary. Also the variance of the latter difference is higher than that of the former. Pooling all the countries, the coefficient of variation, i.e. the ratio between the average and the Standard Deviation, is higher for the high-middle than for the middle-low wage difference. This suggests that most of the difference in the return on education across countries happens at high educational attainment.

The final issue regards the variables that affect education but not wages. Usually, in the literature candidate instruments are related to family background characteristics (Table 5 in Card (1999)). Ideally, I would like to have parents' characteristics. EU-SILC contains information only about parents who are interviewed in the survey, which basically means parents and children living together. There are two issues with this indicator: one is the reduced sample size (the match would be valid for less than 15% of the sample, higher for the mothers than for the father, reflecting families' separations) and one is the potential bias in the estimation induced by sample selection, if the least successful children (parents) live with their parents (children). For these reasons the instrument considered in Section 4 is the educational attainment of the partner. For the present paper, it is reassuring that Trostel et al. (2002) use both parents' and partner's educational attainment and find either remarkably similar results or smaller returns with the latter than with the former instrument. All the following analysis is based on partner's education, whose return may be understood as slightly conservative than those from parents' education. Notwithstanding the cross country heterogeneity, the relevance of this instrument is unquestionable as the correspondence between couples with the same educational attainment is high (slightly less than 40% of the working age population). However, for later reference recall that the information about education of the partner is missing for about 35% of the observations.

4 Results

Table 3 shows the return on education using the methods presented in Section 2. The reference population is made of: working age population, conventionally defined as 25–65 years, so to exclude individuals enrolled in universities or retired; years 2004-2012; men, to avoid a sample selection bias due to labour market participation or preferences for part-time jobs.[7] I estimated several model specifications: the benchmark specification presented in the table always conditions on educational achievement, labour market experience and year dummies. Other specifications are declared throughout the section.

To gain flexibility, rather than estimating a single return for schooling, coefficients are specific to educational achievement. This choice is supported by a large literature, documenting nonlinearities in the return to education (for example, Trostel (2005) rejects the null hypothesis of linearity for almost all the countries, including many analyzed in this paper; below I explicitly address this issue).

4.1 OLS-ATE

With the OLS-ATE, the return for the population at average on middle education with respect to low education (for short, 'the middle-low education') is rather small, usually smaller than 25%: larger returns are estimated only in Luxembourg (37.6%) and Portugal (41.6%). The return on high education with respect to middle education (for short, 'the high-middle education') in Sweden is about 15%, in Norway, Denmark and Belgium is smaller than 25%, whereas in all the other countries is above 25%. These results are consistent with evidence from previous section that most of the heterogenity in return is at the high educational attainment. They also confirm the importance of the non-linearities in returns. More formally, I tested the null hypothesis of equality of the returns between the 'middle-low education' and the 'high-middle education': with the only exception of Norway, the null hypothesis is always rejected, at standard confidence levels.

These returns are fairly similar to those estimated in the existing literature. For Austria,

[7] Excluding students, retired and disabled, more than 90% of men work, whereas the share is only 75% for women. Among workers, 95% of men work full time, opposed to less than 70% for women. Nevertheless, in a robustness check presented later I also include women.

Germany, Ireland, Netherlands and Norway, Trostel (2005) estimates a remarkably similar return exploiting a quadratic function in years of education (which however is more restrictive than the model estimated in this paper). In a study on the college wage gap, Brunello et al. (2000a) estimate similar returns in Austria, France and Netherlands, but slightly smaller for Denmark. For Italy, Brunello et al. (2000b) find similar results using a similar specification. For UK, the estimates are remarkably close to appropriate counterparts in Blundell et al. (2005). When a constant return to schooling across educational attainment is imposed, estimated returns are slightly higher but comparable to those estimated in this paper. This is the restriction in Ichino and Winter-Ebmer (1999) for Germany and in Trostel et al. (2002) for some countries reported in 3. Higher gross wage returns are also estimated in Boarini and Strauss (2010), although only for the year 2001.

With the benchmark specification, the OLS-ATE may be biased due to omitted variables that are unobservable to the researcher. To avoid this risk, for UK Blundell et al. (2005) enrich their basic specification (which is similar to that in Table 3) with controls for family background, namely parents' characteristics. EU-SILC does not provide information about parents' characteristics: this is a major limit, because these variables are among the most important to reduce the OLS-ATE bias in Blundell et al. (2005). Family background with the data at hand is the educational level achieved by the partner.[8] However, it is worth emphasizing that Trostel et al. (2002) find similar results using either parents' or partners' educational level. This greatly mitigates the concern of having only one of the two indicators. The returns with these additional controls are lower than those of the benchmark specification, on average by about 4 percentage points for both educational levels. The content of this robustness check is huge in its own and in relation to the existing literature. Family background has a direct explanation power for the wage process (as found in Blundell et al. (2005)). As a direct consequence, it supports the argument in Card (1999), that the OLS-ATE of earning on schooling *and* family background may have a smaller bias than the OLS-ATE of earning on schooling only.[9]

[8] A priori it is impossible to say which of the two characteristics related to family background is more important. For concreteness, define the omitted variable x_j, where $j = 1$ for father's education and $j = 2$ for partner's education. Then -if $x_1' x_2 = 0$- in eq. 1, $\text{bias}_\beta = (S'S)^{-1} S'(\gamma_1 x_1 + \gamma_2 x_2)$. It may be possible that $S'x_1 \geq S'x_2 \geq 0$ but $\gamma_2 \geq \gamma_1 \geq 0$, so that the marginal contribution of x_1 and x_2 to the bias of β is economically (almost) identical. Whether this is the case or not is only an empirical issue that cannot be addressed a priori.

[9] As an additional robustness check, I also control for a larger set of characteristics. In particular, in addition to

4.2 IV-LATE

As argued in the literature (Harmon et al. (2003)), the OLS-ATE may be biased because schooling may be correlated with components that are unobservable to the researcher, like individual ability ($S_i'\alpha_i \neq 0$), or because workers achieve higher education induced by heterogeneous private return ($S_i'(\beta - b_i) \neq 0$). To cure for these biases, following Trostel et al. (2002) I instrument the schooling achievement with the highest educational attainment reached by the partner.

The sub-section is organized as follows. I first show the results, then I critically review drawbacks with this instrument and propose easy to implement solutions. At the outset, it shall be clear that the limitations of IV-LATE jeopardize a fair cross country comparison and therefore a different approach is needed.

Returns from IV-LATE are substantially higher than those from OLS-ATE. A formal Hausman (1978) test rejects most of the time the null hypothesis that the difference between the two estimates is not systematic. This results follows from a larger difference in the middle-low education return than in the high-middle education: on average, IV-LATE is higher than OLS-ATE by more than 30 percentage points for middle education and about 15 percentage points for high education.[10] It follows that, in contrast to the OLS-ATE, using IV-LATE in 2/3 of the countries the retun on education is linear in educational level (formally: the null hypothesis of equal return between the two educational levels is not rejected).

That returns with OLS-ATE are smaller than with IV-LATE is a fairly standard result in the literature. This is the most compelling puzzle in the field as opposed to the expectation that the most able workers achieve higher education, in which case the OLS-ATE return would be higher than that from IV-LATE. Card (1999) offers different explanations to solve the puzzle, including

Blundell et al. (2005), I enriched the set of controls with dummies for: marital status, managerial position, public sector employment. With respect to Blundell et al. (2005), I have no information regarding test undertaken by workers and the type of school they attended. However, for an international comparison, this is not an issue because it is unclear to what extent this information would be comparable across countries (Peracchi (2006)). The returns with these additional controls are lower than those of the benchmark specification, on average by about 5 percentage points for the middle-low return and 10 for the high-middle education.

[10] The return on middle-low education decreases by 3 percentage points in Netherlands, but increases in all the other countries by more than 10 percentage points, reaching a peak of 50 percentage points or more in Czech Republic and Cyprus. The return on high-middle education decreased by 15 percentage points in Luxembourg and Cyprus, between 5 and 10 in Germany, Ireland and Norway and remained more or less constant in Sweden, Denmark, Island and Italy; in the other countries, IV returns are larger than from OLS-ATE.

1) measurement error (Kane et al. (1999)), 2) legitimacy of instruments (Conley et al. (2012)) and 3) identified quantities (Imbens and Angrist (1994)). In this section, for the sake of brevity I focus only on the last issue (additional material is in Appendix I). However, before proceeding it is interesting to notice that measurement error in schooling achievement does not seem the best candidate to reconcile the difference between β_{IV} and β_{OLS} (see Appendix I.I). As for legitimacy of instrument, the joint null hypotheses of exogeneity and monotonicity based on test proposed by Mourifie and Wan (2016) is rejected. For this reason, I exploit an approach based on 'approximate exogeneity', that allows the instrument not being perfectly exogenous (Conley et al. (2012)). With this correction (Table 4), returns based on OLS-ATE are above those from IV-LATE , thus solving one of the most concerning issue in this field (Appendix I.II).

Closely related to this paper, from Section 2, even with an appropriate instrument, IV-LATE could be higher than OLS-ATE because the two estimators identify different quantities. In this application, IV-LATE is the causal effect for workers who proceed with education *because* their partners proceed. This may be described as the population who can afford to study (i.e., low liquidity constrained), who continue with education because are pushed to do so by their eager partners, who attach high value to education. If these individuals were high ability students they would have gone on studying anyway (i.e., always takers in the terminology of Imbens and Angrist (1994); Angrist et al. (1996)); therefore, compliers are likely to be of low ability. In terms of cost-benefit analysis, the return of these individuals is in the 'middle range' (in case of low ability) with respect to cases of low liquidity constrained/high ability individuals -who enjoy the highest return- and with respect to high liquidity constrained/low ability individuals -who enjoy the lowest return. The argument is adapted from Blundell et al. (2005). Most important for an international comparison, this implies that, no matter what instrument is used, IV-LATE does not readily extend to the whole population, unless are verified special circumstances when LATE and ATE coincide. This identification issue makes a comparison across countries based on IV-LATE useless (even if one is willing to focus on a specific subpopulation), because countries in general differ by sociological background or institutions. As a consequence, the same instrument in two different countries affects two different populations of compliers. For example, if in Italy the role of parents or partner is more relevant to shape the educational choice of the individuals than in UK, the family background as instrument will target two different subpopulations, thus the empirical content of the results would be different between the two countries. This well known property has never been noticed in the existing literature concerning different countries. Yet, it is key to interpret the results.

This problem would be inessential if I could flag the population of compliers. This is not possible, although I can describe the distribution of their characteristics (Angrist (2004)). If the characteristics of compliers are distributed similarly across countries, one may hope that the same instrument targets the same compliers in different countries (though notice that this needs not necessarily be true). Certainly, if the distribution change the same instrument targets different compliers in different countries. Not shown (available on the website), I characterized the distribution of compliers, using the approximate exogenous IV-LATE. This exercise empirically supports that the education of the spouse does not identify the same populations of compliers across countries, because the odds ratio for each covariate is larger than 1 in some countries and smaller in others. Therefore, the more general claim is that when the interest is in a cross-country comparison, IV-LATE is not a good solution. In the large literature interested in international comparisons, I am not aware of any such statement. This conclusion holds even if other widely used instruments are employed (Table 4), like policy interventions or quarter of birth (Angrist and Krueger (1991)). In Appendix I.III, I provide a more thorough discussion of these instruments.

4.3 Bounds

Even though IV-LATE is the most used estimator of the return on education, above results confirm that 'finding a suitable instrument is not an easy task' (Blundell et al. (2005)), and when it is found, the 'credibility of underlying assumptions is always a matter of disagreement' (Manski and Pepper (2000)). When these issues are addressed, the quantities that are identified are inappropriate for an international comparison. Therefore, finding an approach that is robust to all these critiques is necessary to compare estimates across countries, like the return on education around the EU.

To this aim, I estimate the bounds proposed by Manski and Pepper (2000, 2009). I exploit directly the MTR-MTS assumption that is suited even for cases where there is no bounded support of the dependent variable of interest and is consistent with the standard Mincer (1974) equation. A first improvement of bounds over other approaches is that the working hypothesis of MTR-MTS are jointly testable. If they hold, the wage is weakly increasing in the schooling achievement. From Table 2, I check whether the monotonicity assumption holds. The (log) wage from middle education (col.(3)) is always higher than (log) wage of low education (col.(1)) and the (log) wage from high education (col.(5)) is always higher than (log) wage of middle education (col.(3)). The wage process is (strictly) increasing in the educational level, thus supporting the hypothesis. This property makes the international comparison discussed below more credible than those available in the literature.

The average return for middle-low education is up to 25–30%, from a minimum of 14% in Sweden to a maximum of 46% in Luxembourg. The average return for the high-middle education is slightly less than 40%, from 16.7% in Sweden to 75% in Portugal. For both educational level, upper bounds are usually closer to OLS-ATE than to IV-LATE. In particular, although usually higher, upper bounds are economically indistinguishable from OLS-ATE. The average difference between the two estimates is about 5 percentage points for middle-low education, up to 17 ($= 0.402 - 0.231$) in Germany and between 3 and 10 in most of the other countries; for the high-middle education, the average difference between the two estimators is 2 percentage points even if in Greece ($0.436 - 0.335 = 10$ percentage points) and Portugal (20 percentage points), the difference is remarkably larger than in other countries.

Bounds suggest that selection on unobservables is relevant in some countries, but not all, as instead one would conclude using IV-LATE. Whether this happens or not must be judged country-by-country. However, the OLS-ATE alone would not suffice to draw correct conclusions because in some countries the selection on unobservables is indeed a feature of the data. Since I do not use instrument, the result does not rely on exclusion restriction and is not limited to the specific population of compliers: both properties explain why the bounds are closer to OLS-ATE than to IV-LATE.

What is the policy content of these results? The small middle-low return is consistent with the intermediate educational level being the highest share of the working age population, and thus the least remunerated, as found in Pischke and von Wachter (2008). The larger high-middle return reflects the relative scarcity of highly educated individuals: accordingly, in countries like Belgium, Ireland or Spain, where the share of highly educated individuals is higher than that of middle educated, the return is smaller. A remarkable exception to this reasoning is Italy where the share of highly educated individual is one of the lowest among countries analyzed in this paper (slightly larger than 10 percent), yet its return is among the lowest (Visco (2014, 2015) has defined this as a 'paradox'). As a benchmark, in Poland, Hungary, Slovenia and Portugal, other countries where the share of this educational share is smaller than 20 percent, the returns are the highest at 45–55%, no matter the strategy.

5 Some robustness checks

Section 4 offers an updated picture of return on education around Europe. It is broadly consistent with the existing literature but a distinctive feature is that these new results refer to the entire population. It is interesting to see whether results hold across different subsamples of interest, tightly defined. Neither the OLS-ATE nor the IV-LATE are very helpful to this aim: the former is biased, the latter targets a population that is unknown. In the spirit of a robustness check, I discuss only the methodology in Manski and Pepper (2000) which is the preferred approach. Differences in the return on education may arise by age classes, pre- and post- the last economic crisis, excluding public sector workers, excluding immigrants and including women.

It shall be clear that results are broadly robust to these checks. Since the impact on return is heterogeneous across countries, the exercise is also revealing of strength and weaknesses of each country with respect to particular aspects of the national labour markets.

By age

The return on education by age class is probably the most interesting from a policy perspective (Figure 1). The parameter should be equal across different generations, unless there is (a form of) labour market segmentation, different knowledge acquired per year of education, different quality of the teaching, etc. The bounds in this case are modified as in Mariotti and Meinecke (2015), to account for differences in schooling attainment across generations.

There are important differences across countries: in some cases the return on education is broadly similar along the wage distribution, in other cases no. For the oldest and the youngest age classes, results tend to be different from the rest of the population. The smaller return for the youngest age cohorts deserves further investigation of the first job (which I could not identify using EU-SILC).

Only in Hungary and (for a large part of the age range) Netherlands, the returns for both educational levels are constant. In some countries, the return is constant in age for only one educational level: for example, in Germany and Denmark the return on high-middle education is broadly constant, apart at the lowest age ranges, whereas the middle-low return decreases from the age range around 30-40 onwards; in France, the the middle-low return is constant over a wide range of the age distribution, but the high-middle return increases with age. In Cyprus, Czech Republic, Island and Slovak Republic it's hard to identify a clear path of the gap for both educational attainments. Clear evidence of segmentation in both educational attainments are estimated in Spain, Greece, Italy and Portugal, where the return on education is decreasing for both educational levels as going from older to younger generations. Weaker evidence of segmentation is estimated in Austria and Slovenia. In UK, there is a reverse-U shape in returns: both the youngest and the oldest generations are at disadvantage with respect to those in the middle range.

The high-middle return is almost always higher than the middle-low return. Some exceptions are found in the youngest generations, where the higher labour market experience compensate the lower educational attainment at early stage of the career.

With respect to the benchmark case, these results imply that 1) in some countries the inter-generational inequality is relevant, 2) no matter what happens to different generations, there in general exists an educational premium across different educational level, but 3) when this rule is broken, younger generations are more at risk.

Pre- vs Post-crisis

To understand the consequences of the economic crisis on the return on education, I split the sample before and after 2008 (Figure 2). For middle-low education, the number of cases where the pre-crisis return is higher than the post-crisis and vice-versa are broadly balanced. The average difference between the two periods is negligible.

For high-middle education, the post-crisis return is lower than pre-crisis by about 2 percentage points, on average. However, there is a huge heterogeneity. Where the return increased during the crisis, the difference between the two periods is generally small (on average 3 percentage points). In contrast, in several countries where the return decreased after the crisis, the difference between the two periods is higher than 5 percentage points, with a large drop in Cyprus and Portugal (15 percentage points).

In an attempt to shed light on this result, I run two separate regressions (not shown) where the pre-/post-crisis hypothesis is tested within each single age cohort: for high-middle education, younger workers suffer more from the crisis than older workers.

Public sector workers

When I exclude public sector workers, the return on education decreases more in the case of middle-low education than in the high-middle education, reflecting that the largest share of public sector workers hold an intermediate educational level and that public sector workers usually enjoy a wage premium, larger at intermediate level (Depalo et al. (2015)). When the focus is on middle-low education, the return excluding public sector workers is lower in all the countries: only in Greece (7 percentage points), Germany (6 percentage points), Spain (5.5 percentage points) and Hungary the difference is sizeable, but otherwise is smaller than 3 percentage points. For high-middle education the average difference with respect to the benchmark is small and not always negative (although in these cases the difference is economically negligible); larger drops when excluding public sector employees are estimated in Greece (about 8 percentage points) and in Spain, Italy and Portugal (about 7 percentage points).

There are three important results to emphasize with this subsample: first, the return on education is still increasing in the educational attainment; second, the non linearity (i.e., the difference between the high-middle and middle-low return) is larger than in the benchmark case (larger than 3.5 percentage points), apart from few exceptions where the opposite happens; third, among these exceptions are Spain, Greece, Italy, Portugal, where public sector workers enjoy a higher premium than in other countries (Depalo et al. (2015)).

Excluding immigrants

When I exclude immigrant workers, the correction with respect to the benchmark specification is generally negligible for the high-middle education and larger for the middle-low education. In Luxembourg the non-linearity decreases by 6 percentage points with respect to the benchmark, as a consequence of the high share of immigrants with high education who also enjoy high earnings in this country; in Austria and Latvia, it decreases by more than 2-3 percentage points. At the opposite side, in Spain and Greece the non-linearity increases by more than 2 percentage points.

At high education, the difference between this sample and the benchmark is almost always negligible in economic terms, apart from Luxembourg, where the high-middle education return decreases by 13 percentage points if I exclude immigrants, therefore for these immigrants return on education is higher than for natives. An interesting finding is the correspondence between return on high-middle education and distribution of education across different sub-populations: on average, the higher the share of immigrants with high education with respect to natives, the smaller the drop in return of this educational level. Highly educated immigrants are not at disadvantages with respect to natives.

Focusing on middle-low education, the difference in return with respect to the benchmark specification is larger.[11] In this case, the return on middle-low education increases if the share of immigrants with middle education excluding immigrant is larger than that of natives with middle

[11] In Luxembourg, the middle-low return decreases by 20 percentage points; it also decreases in Austria (4 percentage points), Belgium (2 percentage points), Norway, France and UK (1–1.5 percentage points). In contrast, when the return increases the difference is smaller: in Portugal and Spain it is 2 percentage points, in Ireland, Slovenia, Greece is about 1, in Italy and Germany is between .5-1 percentage points. In Luxembourg middle educated immigrants are 30 percentage points less than in the native population, in UK 17, in France 16, Norway and Austria 14, Belgium 6; coherently, in Portugal middle educate immigrants are more than in the native population by 16 percentage points, Spain 10, Greece 2.5.

education. Put otherwise, non high-skilled immigrants are at disadvantage with respect to natives with similar characteristics.

Including women

For women, the decision about labour market participation might be more complex than purely economic considerations, as shown by the lower participation rates with respect to men. To limit the risk of discrimination linked to part-time versus full-time employees, I discuss the results excluding part time workers (Moffitt (1984)). This decision should prevent from the estimation of spurious relations.

In about half of the countries, the middle-low return increases with respect to the benchmark and in about half decreases. Instead, the difference in return of high-middle education is large with respect to the benchmark (with remarkable cross countries distinctions), almost always in favour of men: in most countries, the difference in return is about 5 percentage points, whereas in Greece (about 10 percentage points), Hungary, Island and Italy (8 percentage points) is higher. Only in Portugal, Czech Republic and Spain I do not estimate a significant gender wage gap, thus women are not at disadvantage with respect to men.

To check the source of this wage gap, I investigated the sector of employment of women with high education: women are the majority of workers in low paid sectors, in particular hotel, retail and education. A possible explanation is the degree of substitutability between male and female workers (De Giorgi et al. (2013)). The investigation of the sorting mechanism is an intriguing question that is beyond the scope of this paper and the I leave for future research.

6 Going beyond bounds

Once it is clarified that bounds can solve the issues that jeopardize a comparison across countries with standard estimators, it is interestring to go a step further beyond the simple return on education (Section 6.1). There are two important issues that are worth exploring: one is the information contained in bounds beyond the simple return on education (Section 6.1); one is the relation between bounds and other estimators (Section 6.2).

6.1 Going beyond bounds

To go beyond the plain return on education, it is interesting to look at the marginal contribution of each component to the identification of bounds. The representation proposed in this section provides further information regarding the differences in returns across countries. To shorten the notation, 'L' is used for 'Low'; 'M' is used for 'Medium'; 'H' is used for 'High' educational level. For the middle-low bounds, the contributions to identification are equal to

$$\begin{aligned} \beta_{ML} &= \{E[y|z=M]-E[y|z=L]\} + P[z=H] \quad \{E[y|z=H]-E[y|z=M] \\ &= \underbrace{\{P[z=L]+P[z=M]\}}_{\text{Pop.Effect}} \underbrace{\{E[y|z=M]-E[y|z=L]\}}_{\text{Ass.Effect}} + \underbrace{P[z=H]}_{\text{Pop.Effect}} \underbrace{\{E[y|z=H]-E[y|z=L]\}}_{\text{Ass.Effect}} \end{aligned} \tag{2}$$

whereas for the high-middle bounds, the contributions are

$$\begin{aligned} \beta_{HM} &= P[z=L] \quad \{E[y|z=M]-E[y|z=L]\} + \quad \{E[y|z=H]-E[y|z=M \\ &= \underbrace{P[z=L]}_{\text{Pop.Effect}} \underbrace{\{E[y|z=H]-E[y|z=L]\}}_{\text{Ass.Effect}} + \underbrace{\{P[z=M]+P[z=H]\}}_{\text{Pop.Effect}} \underbrace{\{E[y|z=H]-E[y|z=M}_{\text{Ass.Effect}} \end{aligned} \tag{3}$$

where the 'Assumption effect' is the gain associated to the imposed hypothesis and the 'population effect' is the population to which it applies. The first line of equations 2 is very useful to clarify that in β_{ML} I observe $E[y|z=M]$ and $E[y|z=L]$, but I can only conjecture that the wage for the population of high education is on average no smaller than $E[y|z=M]$; this restriction applies only to unobserved outcomes, i.e. for individuals with highest educational level ($P[z=H]$). The same reasoning applies to β_{HM}.

A related exercise derives the difference between the two educational levels as

$$\begin{aligned} \Delta = \beta_{HM}-\beta_{ML} &= \underbrace{\{P[z=L]+2\,P[z=M]+P[z=H]\}}_{\text{Pop.Effect}} \underbrace{\{E[y|z=L]-E[y|z=M]\}}_{\text{Ass.Effect}} \\ &+ \underbrace{\{P[z=L]+P[z=M]\}}_{\text{Pop.Effect}} \underbrace{\{E[y|z=H]-E[y|z=L]\}}_{\text{Ass.Effect}} \end{aligned}$$

$$
\begin{aligned}
= \quad & \underbrace{\{P[z=L]+P[z=M]\}}_{\text{Pop.Effect}}\underbrace{\{E[y|z=H]-E[y|z=M]\}}_{\text{Ass.Effect}} \\
+ \quad & \underbrace{\{P[z=M]+P[z=H]\}}_{\text{Pop.Effect}}\underbrace{\{E[y|z=L]-E[y|z=M]\}}_{\text{Ass.Effect}},
\end{aligned} \tag{4}
$$

where $E[y|z=L]-E[y|z=M] \leq 0$ and $E[y|z=H]-E[y|z=L] \geq 0$ or $E[y|z=L]-E[y|z=M] \leq 0$ and $E[y|z=H]-E[y|z=M] \geq 0$ for the technique to be employed (coherent with the economic model in Mincer (1974) and the working hypothesis in Manski and Pepper (2000)). Therefore, Δ increases with $[E[y|z=H]-E[y|z=M]]$, but decreases as the return on middle education with respect to low education becomes higher ($[E[y|z=L]-E[y|z=M]] < 0$). The population effect amplifies the role of assumptions.

This representation gives the opportunity to inspect closer the information provided by bounds. If one estimates that $\Delta < 0$, or concavity in return on education, it may simply signal a limit in the data availability. Consider the following extreme example, where $E[y|z=L] < E[y|z=M] < E[y|z=H]$ with populations $P[z=M]=P[z=L] \cong 0$ and $P[z=H] \cong 1$. Then

$$
P[z=H] \cong 1 \quad \Rightarrow \qquad \Delta \cong -([E[y|z=M]-E[y|z=L]) < 0 \ , \tag{5}
$$

because $\beta_{ML} = \{E[y|z=H]-E[y|z=L]\}$ and $\beta_{HM} = \{E[y|z=H]-E[y|z=M]\}$. This would only reflect that the assumption regarding high-middle return is imposed to those who did not reach the highest education. For some researchers, this might seem an unpalatable property of bounds. In fact, this is only a fair representation of what the data tell: if not much information is in the data, not much results can be derived - unless one is willing to impose not always credible assumptions. Whilst bounds make explicit what is observed and what not and what are the assumptions, standard approaches would simply be silent on this issue, that would be artificially overcome thank to untestable hypotheses. Looking closer at each component reveals much of the reasons why the return on education is so different across countries.

The high-medium return is usually larger than the medium-low return, on average by 10 percentage points (up to 37 in Portugal). Within the assumption effect, the component $E[y|z = H] - E[y|z = M]$ is larger than $E[y|z = M] - E[y|z = L]$ and with a larger variability, thus confirming that much of the cross countries heterogeneity is in the highest educational level. Only in Germany, Norway and Austria the high-medium bound is smaller than the medium-low gain, although by a very small amount. Employing equation 4 this result is due to the uncertainty about the high-middle return (i.e., $E[y|z = H] - E[y|z = M]$), because the population effect $P[z = L] + P[z = M] < P[z = M] + P[z = H]$. In these countries, $(1 - P[z = L])$ is among the largest (more than 95% in Germany, 90% in Austria and 88% in Norway) and with respect to others countries $(1 - P[z = H])$ is lower (for example, in Germany and Czech Republic, $(1 - P[z = L])$ is similar, but $(1 - P[z = H])$ in Germany is only 54%, compared to about 85% in Czech Republic; in norway is 62%; only in Austria is larger at 78%). This shows the data limitation argument at work.

A distinct important question is whether more assumptions could be imposed. For example, Okumura and Usui (2014) impose log concavity in the return on education -which narrows the Manski and Pepper (2000) bounds without using instrument. Results from Table 3 suggest that this hypothesis would not be appropriate with these data.

6.2 On the relation between various estimators

The difference in magnitude across estimators suggests to take a step further, to explain where and why they differ - so as to extract as much information as possible. One possible representation is $\Delta_{UPPER-IV} = \Delta_{UPPER-OLS} + \Delta_{OLS-IV}$, where Δ_{i-j} is the difference between the estimator i and j. The OLS-ATE estimates return on education for the population on average, but is not robust against possible endogeneity of education. The IV-LATE is robust to endogeneity of schooling, but estimates the return to schooling only for compliers. Bounds of Manski and Pepper (2000) apply to the whole population, as defined by conditioning covariates and is robust to endogeneity. In this respect, $\Delta_{UPPER-IV}$ may be interpreted as the effect of compliers as opposed to the rest of the population, $\Delta_{UPPER-OLS}$ as the effect of endogeneity in the population and Δ_{OLS-IV} as the 'identification difference' between ATE and LATE. These quantities may be useful to test various hypotheses, throug null hypotheses involving $\Delta_{i-j} = 0$. For example, using the correction in Conley et al. (2012) (see Appendix I.II for further details), none of the quantities involved is equal to zero, for both educational level. Further, for middle-low return $\Delta_{UPPER-OLS} > \Delta_{OLS-IV}$, thus suggesting that endogeneity is an issue for the whole population. Only exceptions are Austria, Spain, Greece, Italy, Latvia, Portugal and Slovak Republic, where Δ_{OLS-IV} is slightly larger than $\Delta_{UPPER-OLS}$, thus not only is endogeneity relevant, but also heterogeneity is far more critical than in other countries. In contrast, for high-middle return $\Delta_{UPPER-IV} > \Delta_{OLS-IV} > \Delta_{UPPER-OLS}$ in 15 countries, thus suggesting that endogeneity matters but also that heterogeneity of the population is substantial. The large heterogeneity in $\Delta_{IV-UPPER}$ strongly supports that the subpopulation of compliers is much different from country to country. Therefore, having a clear insight about the population of compliers is a necessary step toward a better understanding of the population to which the estimated parameter applies, overall at the highest educational level.

7 Concluding remarks

Wage return on education is a key policy indicator within and between countries. Not surprisingly, the estimate of this parameter received much attention in the literature. Two competing estimators are invariably used: the OLS or the IV. The main point of this paper is that none of the two is appropriate for an international comparison, even though no other solution has been proposed before.

The OLS-ATE recovers the return on education for the population on average, but is not robust to possible endogeneity of schooling achievement: as a consequence it is inconsistent. The IV-LATE is robust to endogeneity, but recovers the return on education only for compliers, that cannot be identified from observable data: as a consequence it is consistent, but compares different populations. To avoid both limits, I use set identification, that is robust to possible endogeneity of schooling and is valid for the population as identified by the data.

The technical points of this paper are far more general than it may appear at first glance, because cast doubt on previous cross-country estimates based on IV-LATE in any field. As a distinct contribution to the existing literature on bounds, I also derived the contribution of each single component to identification, which is informative about *i)* the specific differences across countries and *ii)* the main sources of uncertainty in the estimated returns. I also derived a statistic regarding why estimators differ (e.g., endogeneity versus heterogeneity).

Taking advantage of these features, in this paper I update the figures concerning the return on education across countries belonging to the European Union. Return on education are quite different across countries. A useful interpretation of results is in terms of demand/supply schedule. The small middle-low return is consistent with the intermediate educational level being the highest share of the working age population, and thus the least remunerated. The larger high-middle return reflects the relative scarcity of highly educated individuals. The only exception to this argument is Italy where both the share of highly educated individual and return on highest education are among the lowest across countries analyzed in this paper (Visco (2014, 2015) has defined this as a 'paradox'). I leave for further research a better understanding of this paradox.

References

Angrist, J. and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice*. *The Economic Journal 114*(494), C52–C83.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables (Disc: p456-472). *Journal of the American Statistical Association 91*, 444–455.

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

Battistin, E., M. De Nadai, and B. Sianesi (2014). Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics 181*(2), 136 – 150.

Battistin, E. and B. Sianesi (2011, May). Misclassified Treatment Status and Treatment Effects: An Application to Returns to Education in the United Kingdom. *The Review of Economics and Statistics 93*(2), 495–509.

Black, D. A., M. C. Berger, and F. A. Scott (2000). Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association 95*(451), 739–748.

Blundell, R., L. Dearden, and B. Sianesi (2005). Evaluating the effect of education on earnings: models, methods and results from the national child development survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 168*(3), 473–512.

Blundell, R. and M. C. Dias (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources 44*(3), 565–640.

Boarini, R. and H. Strauss (2010). What is the private return to tertiary education? new evidence from 21 oecd countries. *OECD Journal: Economic Studies 10*, 1–25.

Brunello, G., S. Comi, and C. Lucifora (2000a, December). The College Wage Gap in 10 European Countries: Evidence from Two Cohorts. IZA Discussion Papers 228, Institute for the Study of Labor (IZA).

Brunello, G., S. Comi, and C. Lucifora (2000b, March). The Returns to Education in Italy: A New Look at the Evidence. IZA Discussion Papers 130, Institute for the Study of Labor (IZA).

Cappellari, L. and K. Tatsiramos (2015). With a little help from my friends? quality of social networks, job finding and job match quality. *European Economic Review 78*(0), 55 – 75.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*(2), 667–737.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012, February). Plausibly Exogenous. *The Review of Economics and Statistics 94*(1), 260–272.

Conti, G., J. Heckman, and S. Urzua (2010, May). The Education-Health Gradient. *American Economic Review 100*(2), 234–38.

De Giorgi, G., M. Paccagnella, and M. Pellizzari (2013, June). Gender complementarities in the labor market. Questioni di Economia e Finanza (Occasional Papers) 183, Bank of Italy, Economic Research and International Relations Area.

Depalo, D., R. Giordano, and E. Papapetrou (2015). Publicprivate wage differentials in euro-area countries: evidence from quantile decomposition analysis. *Empirical Economics 49*(3), 985–1015.

European Commission (2015). Compulsory education in europe 2014/15.

Flores, C. A. and A. Flores-Lagunes (2013, October). Partial Identification of Local Average Treatment Effects With an Invalid Instrument. *Journal of Business & Economic Statistics 31*(4), 534–545.

Ginther, D. K. (2000, February). Alternative Estimates of the Effect of Schooling on Earnings. *The Review of Economics and Statistics 82*(1), 103–116.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review 73*(C), 103–130.

Hanushek, E. A. and L. Woessmann (2011, May/June). *The Economics of International Differences in Educational Achievement*, Volume 3 of *Handbook of the Economics of Education*, Chapter 2, pp. 89–200. Elsevier.

Harmon, C., H. Oosterbeek, and I. Walker (2003). The returns to education: Microeconomics. *Journal of Economic Surveys 17*(2), 115–156.

Hausman, J. A. (1978, November). Specification Tests in Econometrics. *Econometrica 46*(6), 1251–71.

Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *The Journal of Human Resources 32*(3), 441–462.

Ho, K. and A. M. Rosen (2015, October). Partial identification in applied research: Benefits and challenges. Working Paper 21641, National Bureau of Economic Research.

Ichino, A. and R. Winter-Ebmer (1999, April). Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments. *European Economic Review 43*(4-6), 889–901.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–75.

Kamhöfer, D. A. and H. Schmitz (2015). Reanalyzing zero returns to education in germany. *Journal of Applied Econometrics*, n/a–n/a.

Kane, T. J., C. E. Rouse, and D. Staiger (1999, July). Estimating Returns to Schooling When Schooling is Misreported. NBER Working Papers 7235, National Bureau of Economic Research, Inc.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica 83*(5), 2043–2063.

Kline, P. (2011). The Impact of Juvenile Curfew Laws on Arrests of Youth and Adults. *American Law and Economics Review 14*(1), 44–67.

Lance, L. (2011). Chapter 2 - nonproduction benefits of education: Crime, health, and good citizenship. In S. M. Eric A. Hanushek and L. Woessmann (Eds.), *Handbook of The Economics of Education*, Volume 4 of *Handbook of the Economics of Education*, pp. 183 – 282. Elsevier.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association 90*(431), 1112–1121.

Little, R. J. A. and D. B. Rubin (1986). *Statistical Analysis with Missing Data.* New York, NY, USA: John Wiley & Sons, Inc.

Lochner, L. and E. Moretti (2004, March). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review 94*(1), 155–189.

Manski, C. F. (1990, May). Nonparametric Bounds on Treatment Effects. *American Economic Review 80*(2), 319–23.

Manski, C. F. (2011, 08). Policy Analysis with Incredible Certitude. *Economic Journal 121*(554), F261–F289.

Manski, C. F. and J. V. Pepper (2000, July). Monotone Instrumental Variables, with an Application to the Returns to Schooling. *Econometrica 68*(4), 997–1012.

Manski, C. F. and J. V. Pepper (2009, 01). More on monotone instrumental variables. *Econometrics Journal 12*(s1), S200–S216.

Mariotti, M. and J. Meinecke (2015). Partial Identification and Bound Estimation of the Average Treatment Effect of Education on Earnings for South Africa. *Oxford Bulletin of Economics and Statistics 77*(2), 210–233.

Mincer, J. A. (1974, August). Introduction to schooling, experience, and earnings. In *Schooling, Experience, and Earnings*, NBER Chapters, pp. 1–4. National Bureau of Economic Research, Inc.

Moffitt, R. (1984, October). The Estimation of a Joint Wage-Hours Labor Supply Model. *Journal of Labor Economics 2*(4), 550–66.

Mourifie, I. and Y. Wan (2016). Testing local average treatment effect assumptions. *Review of Economics and Statistics na*, n/a–n/a.

Nevo, A. and A. M. Rosen (2012, August). Identification With Imperfect Instruments. *The Review of Economics and Statistics 94*(3), 659–671.

Okumura, T. and E. Usui (2014). Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling. *Quantitative Economics 5*(1), 175–194.

Peracchi, F. (2006, May/June). *Educational Wage Premia and the Distribution of Earnings: An International Perspective*, Volume 1 of *Handbook of the Economics of Education*, Chapter 5, pp. 189–254. Elsevier.

Pischke, J.-S. and T. von Wachter (2008, August). Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation. *The Review of Economics and Statistics 90*(3), 592–598.

Psacharopoulos, G. (1985). Returns to education: A further international update and implications. *Journal of Human Resources 20*(4), 583–604.

Sianesi, B. and J. Van Reenen (2003). The returns to education: Macroeconomics. *Journal of Economic Surveys 17*(2), 157–200.

Trostel, P. (2005). Nonlinearity in the return to education. *Journal of Applied Economics VIII*(1), 191 – 202.

Trostel, P., I. Walker, and P. Woolley (2002). Estimates of the economic return to schooling for 28 countries. *Labour Economics 9*(1), 1 – 16.

Visco, I. (2014). *Investire in conoscenza. Crescita economica e competenze per il XXI secolo.* Contemporanea. Il Mulino. In Italian.

Visco, I. (2015). Capitale umano e crescita. Invited speech at "Universita' Cattolica del Sacro Cuore-Facolta' di Economia, sede di Roma" – IL FUTURO NELL'ECONOMIA; In Italian.

von Simson, K. (2015). Explaining upper secondary school dropout: new evidence on the role of local labor markets. *Empirical Economics 48*(4), 1419–1444.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources 50*(2), 420–445.

Table 1: A comparison across different estimators

Method	Robust to α	Refers to whole Pop.	Type of Ident.	Credibility
OLS	X	V	Point	?
IV	V	X	Point	?
Bounds	V	V	Set	Yes

Note: X=No; Y=Yes; ?=Maybe.

Table 2: Descriptive statistics

Country	Low		Middle		High	
	log(Wage)	Share	log(Wage)	Share	log(Wage)	Share
	(1)	(2)	(3)	(4)	(5)	(6)
AT	2.18676	9.3	2.42483	69.1	2.66886	21.6
BE	2.34406	20.5	2.42489	39.3	2.62894	40.1
CY	2.27954	19.9	2.43283	40.6	2.80356	39.5
CZ	0.99438	4.1	1.23918	79.3	1.62290	16.6
DE	2.46426	4.1	2.70336	50.3	3.05993	45.6
DK	3.06203	16.1	3.17259	50.4	3.41150	33.5
EE	0.95435	10.7	1.08831	66.3	1.38891	23.0
ES	1.92701	43.0	2.11897	23.5	2.35337	33.5
FR	2.27729	18.7	2.36861	50.0	2.67343	31.3
GR	1.77604	28.0	1.93150	43.2	2.32401	28.8
HU	0.79295	10.8	1.05570	69.6	1.70398	19.6
IE	2.55691	30.1	2.67873	29.1	2.93160	40.8
IS	2.67858	23.6	2.87576	48.8	3.16394	27.7
IT	2.10139	40.3	2.28824	47.0	2.56337	12.6
LT	0.70461	6.6	0.81858	64.2	1.25481	29.2
LU	2.53931	32.3	2.88949	36.0	3.24226	31.7
LV	0.80202	14.6	0.94964	62.9	1.37723	22.5
NL	2.88901	19.2	3.04427	41.9	3.38436	38.9
NO	3.02624	12.1	3.21919	50.2	3.44941	37.7
PL	0.70944	6.7	0.90566	73.3	1.37014	20.0
PT	1.47261	72.7	1.79034	16.2	2.30333	11.1
SE	2.43150	12.4	2.52752	56.9	2.68301	30.6
SI	1.47458	14.7	1.67586	65.8	2.17674	19.5
SK	0.96607	2.5	1.15606	76.1	1.52151	21.5
UK	2.51086	12.2	2.74354	48.0	3.06811	39.8

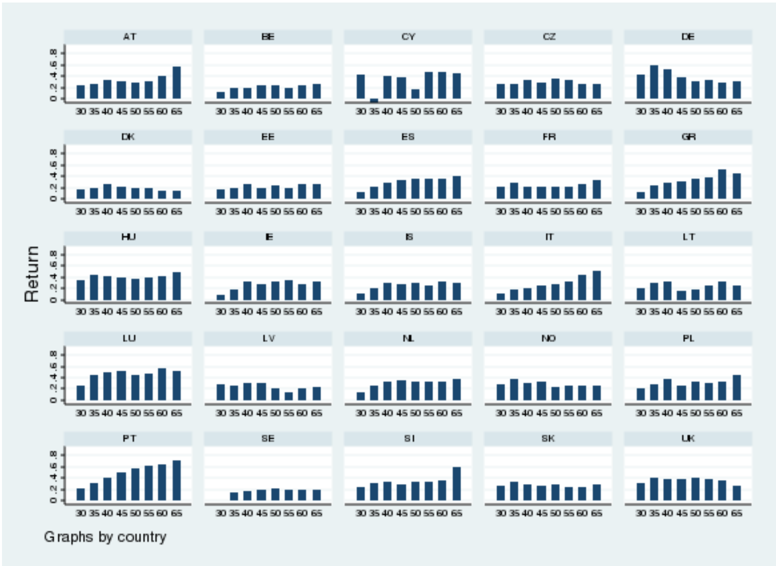
Table 3: Estimation by educational level

	OLS		LATE		Bounds	
	Middle-Low	High-Middle	Middle-Low	High-Middle	Middle-Low	High-Middle
AT	0.237 ***	0.268 ***	0.718 ***	0.490 ***	0.291 ***	0.266 ***
BE	0.134 ***	0.238 ***	0.259 ***	0.274 ***	0.163 ***	0.221 ***
CY	0.185 **	0.415 ***	0.746 **	0.255	0.300 ***	0.401 ***
CZ	0.215 ***	0.383 ***	0.739 ***	0.505 ***	0.309 ***	0.394 ***
DE	0.231 ***	0.341 ***	0.484 ***	0.242 ***	0.402 ***	0.366 ***
DK	0.116 ***	0.231 ***	0.509 ***	0.211 ***	0.191 ***	0.257 ***
EE	0.152 ***	0.282 ***	0.552 ***	0.460 ***	0.203 ***	0.315 ***
ES	0.237 ***	0.275 ***	0.350 ***	0.485 ***	0.270 ***	0.317 ***
FR	0.130 ***	0.389 ***	0.460 ***	0.471 ***	0.187 ***	0.322 ***
GR	0.215 ***	0.335 ***	0.454 ***	0.375 ***	0.269 ***	0.436 ***
HU	0.258 ***	0.653 ***	0.568 ***	0.820 ***	0.390 ***	0.677 ***
IE	0.187 ***	0.256 ***	0.395 ***	0.169 ***	0.225 ***	0.290 ***
IS	0.164 ***	0.306 ***	0.304 ***	0.292 ***	0.277 ***	0.335 ***
IT	0.245 ***	0.315 ***	0.462 ***	0.305 ***	0.222 ***	0.350 ***
LT	0.130 ***	0.429 ***	0.532 ***	0.555 ***	0.241 ***	0.444 ***
LU	0.376 ***	0.459 ***	0.802 ***	0.287 ***	0.462 ***	0.466 ***
LV	0.176 ***	0.406 ***	0.520 ***	0.614 ***	0.244 ***	0.449 ***
NL	0.173 ***	0.375 ***	0.140	0.512 ***	0.288 ***	0.370 ***
NO	0.221 ***	0.212 ***	0.615 ***	0.146 **	0.280 ***	0.254 ***
PL	0.215 ***	0.534 ***	0.616 ***	0.702 ***	0.289 ***	0.478 ***
PT	0.416 ***	0.539 ***	0.669 ***	0.575 ***	0.375 ***	0.744 ***
SE	0.108 ***	0.163 ***	0.539 ***	0.135 ***	0.144 ***	0.167 ***
SI	0.233 ***	0.520 ***	0.463 ***	0.670 ***	0.299 ***	0.530 ***
SK	0.199 ***	0.355 ***	0.659 ***	0.486 ***	0.268 ***	0.370 ***
UK	0.247 ***	0.344 ***	0.430 ***	0.517 ***	0.362 ***	0.353 ***

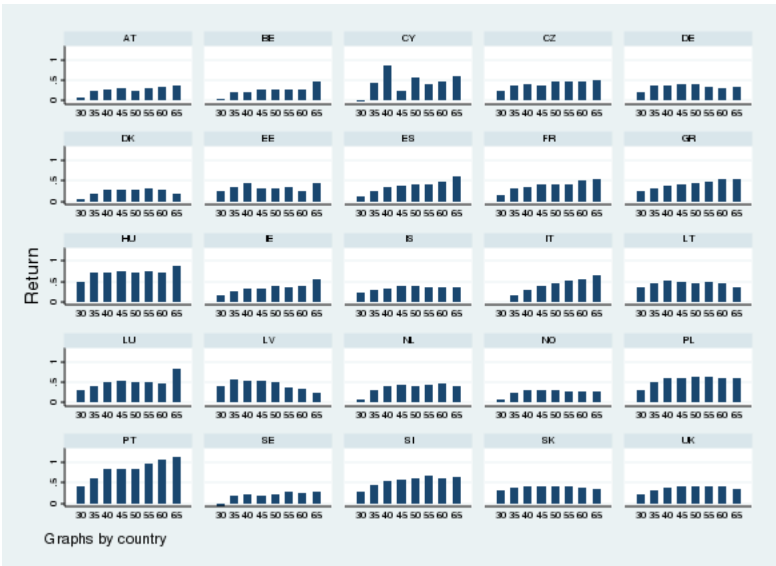
Table 4: Various IV estimation by educational level. The point estimates exploiting the 'Palusible exogenous' approach in Conley et al. (2012) are based on γ derived from a regression of y on benchmark covariates plus the educational attainment achieved by the partner (the complete set of combinations is on my website).

	Low			High		
Instrument:	Partner	Partner	Semester	Partner	Partner	Semester
Method:	Standard	Plausible ex.	Standard	Standard	Plausible ex.	Standard
AT	0.718 ***	0.166 ***	0.261 ***	0.490 ***	0.215 ***	0.250 ***
BE	0.259 ***	0.114 **	0.154 ***	0.274 ***	0.228 ***	0.212 ***
CY	0.746 **	0.130	0.025	0.255	0.500 ***	0.510 ***
CZ	0.739 ***	0.166 **	0.229 ***	0.505 ***	0.344 ***	0.383 ***
DE	0.484 ***	0.161 *		0.242 ***	0.344 ***	
DK	0.509 ***	0.098	0.124 ***	0.211 ***	0.216 ***	0.209 ***
EE	0.552 ***	0.114	0.154 ***	0.460 ***	0.230 ***	0.286 ***
ES	0.350 ***	0.197 ***	0.254 ***	0.485 ***	0.254 ***	0.250 ***
FR	0.460 ***	0.093 *	0.146 ***	0.471 ***	0.342 ***	0.356 ***
GR	0.454 ***	0.160 ***	0.236 ***	0.375 ***	0.286 ***	0.300 ***
HU	0.568 ***	0.170 ***	0.250 ***	0.820 ***	0.612 ***	0.669 ***
IE	0.395 ***	0.171 ***		0.169 ***	0.254 ***	
IS	0.304 ***	0.161	0.214 ***	0.292 ***	0.261 ***	0.276 ***
IT	0.462 ***	0.212 ***	0.251 ***	0.305 ***	0.321 ***	0.328 ***
LT	0.532 ***	0.068	0.132 ***	0.555 ***	0.377 ***	0.382 ***
LU	0.802 ***	0.330 ***	0.345 ***	0.287 ***	0.422 ***	0.483 ***
LV	0.520 ***	0.089	0.211 ***	0.614 ***	0.373 ***	0.376 ***
NL	0.140	0.156		0.512 ***	0.348 ***	
NO	0.615 ***	0.178	0.227 ***	0.146 **	0.202 ***	0.228 ***
PL	0.616 ***	0.162 ***	0.213 ***	0.702 ***	0.505 ***	0.539 ***
PT	0.669 ***	0.387 ***	0.474 ***	0.575 ***	0.499 ***	0.461 ***
SE	0.539 ***	0.094	0.133 ***	0.135 ***	0.138 ***	0.155 ***
SI	0.463 ***	0.213 ***		0.670 ***	0.476 ***	
SK	0.659 ***	0.129	0.201 ***	0.486 ***	0.306 ***	0.334 ***
UK	0.430 ***	0.214 ***		0.517 ***	0.305 ***	

Figure 1: Estimates by educational attainment and age - Manski and Pepper (2000)



(a) Middle-Low



(b) High-Middle

Figure 2: Estimates by educational attainment; various subsmaples - Manski and Pepper (2000)

(a) Middle-Low

(b) High-Middle

I Appendix: More on IV

I.I Measurement errors

Much of the literature on schooling reach the conclusion that the indicator is fairly reliable. An advantage of EU-SILC is the broad classification of the schooling achievement, which limits the possible measurement error. To fix ideas, I check whether this claim is true or not exploiting the classical measurement error framework.[12] Start from eq. 1, if I am interested in $y = S\beta + \epsilon$ but observe $\tilde{S} = S + v$, then the estimated $\hat{\beta_{OLS}} = \beta\frac{\sigma_S^2}{\sigma_S^2+\sigma_v^2}$, with σ_j^2 the variance of j. From this formula, it is possible to evaluate how large should be σ_v^2 for β_{OLS} to be attenuated only by measurement error. In this application, for the OLS-ATE to be coherent with the mis-classification of educational achievement, the variance of the error should be as large as 50% of the education for middle education and smaller for the high educational level (more than 10% for some countries, but usually larger than 20%). This large error is possible but highly implausible, as also concluded for example by Battistin and Sianesi (2011); Battistin et al. (2014) for UK.

Measurement error in schooling achievement does not seem the best candidate to reconcile the difference between β_{IV} and β_{OLS}.

I.II Legitimacy of instrument

The validity of family background as instrument has been questioned in previous papers. If family background is correlated with the outcome of interest, IV-LATE may be even more upward biased than the OLS-ATE from the benchmark specification. As a consequence, despite the faith in this instrument, papers using parents' educational achievement may largely *over*estimate the return on education. In this paper, the instrument is the education of the partner, for which exogeneity is a critical issue that would be easily violated if the partner provides job matches with higher wages for his spouse (i.e., networks; see for example Cappellari and Tatsiramos (2015)).[13] Since I have

[12]The non-classical measurement error setup as in Black et al. (2000) is more appropriate, but does not add much to the scope of this section.

[13]There are at least two other reasons of concern with the education of the partner, since it is missing for those who currently do not have a partner: first and obvious, the current partner may not be the same of the time when the educational choice was made (this is closely related to the relevance of instrument; if partner is met through colleagues or is a colleague, also exogeneity is likely to fail); second, the subsample of those living in a consensual

already checked that the variable has a direct explanation power for the wage process, like for parents' educational level one should be cautious when using this instrument (as emphasized by Trostel et al. (2002)). The test on exogeneity and monotonicity proposed by Mourifie and Wan (2016) confirms this result.

Several solutions have been proposed in the literature to correct for imperfect instruments. Some are based on bounds, like Nevo and Rosen (2012), Chernozhukov et al. (2013), Flores and Flores-Lagunes (2013). To keep things as simple as possible, I apply here the method suggested in Conley et al. (2012), that is suited for cases when the validity of instruments is debatable, i.e. only 'plausibly exogenous' or 'approximately exogenous' in their terminology. Using notation of equation 1, this implies that $y_i = \beta\, S_i + \gamma Z_i + \epsilon_i$: under optimal IV framework $\gamma = 0$, whilst if the instrument is only 'approximate exogenous' $\gamma \neq 0$. The intuition is that if I knew the true value of γ, I could purge its effect on y and estimate the model on a newly defined dependent variable $\tilde{y}_i = y_i - \gamma Z_i$. In fact, I don't know γ, but I can impose $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and estimate the corresponding $\beta \equiv \beta(\gamma)$, where the notation makes clear that β depends on γ. I follow this approach in Table 4, where the standard IV and the approximate exogeneity correction are reported. The γ shown in the table is from an OLS from the benchmark regression, controlling only for labour market experience, schooling and year dummies. Results hold over a wide range of γ (the complete set is available on the website). The correction is substantial and larger for the middle-low level (on average by 35 percentage points) than for the high-middle (on average, slightly less than 10 points).[14] Focusing on these new estimates, in all the countries the high-middle return is again above the middle-low return, on average by 17 percentage points. In particular I can distinguish among three groups. In Norway, Sweden and Austria, the return is similar across the two grades (i.e., constant return on education would be a good fit of the result), in some countries the high-middle return is larger

union may suffer from sample selection, if not having a partner is due to preferences for professional life over private life.

[14]The largest difference for middle-low return estimates is in Cyprus, where the estimates coherent with the implausible exogenous estimate is now about 13% compared to 75% with the standard approach; apart from Netherlands, where the correction is inessential, even in countries where the difference is smaller the new estimates are much different, by at least 15 percentage points (e.g., in Island, Belgium and Spain). As for the high-middle return, the correction is of much smaller magnitude: only in 6 countries the new estimates are smaller for an amount larger than 20 percentage points (Austria, Latvia, Spain, Estonia, UK, Hungary); in some countries, the new estimates are higher than those previously estimated, although usually by a small amount; only in Germany (10%), Luxembourg (slightly less than 14%) and Cyprus (about 25%), the new estimates are higher by an economically significant amount.

than the middle-low return by about 10 percentage points (among these countries there are Spain, Ireland, Italy, Portugal and Greece) and in the remaining countries the difference in the returns is 20 percentage points (among these countries there are Germany, France and East-European countries) up to 45 percentage points in Hungary.

The most important result is that the OLS-ATE return is now above the IV-LATE return. This result is particularly relevant because it solves the most compelling puzzle of this field: as suggested in Card (1999), an important drawback is that the instruments used in the literature might not be truly exogenous. When this pitfall is cured, even with very simple strategies, this specific puzzle is solved.

I.III Additional instruments

Several papers exploit policy changes or quarter of birth to identify the return on education. As for policy changes, to be comparable across EU-countries, the same policy should be introduced in all the countries at the same time, which so far has never happened in Europe. Second best solutions would be *some* policy changes taking place simultaneously across countries or the same policy being introduced across countries *at some point* (like an 'event study' framework; Kline (2011)). Even this second best solutions have never happened in Europe during the period analyzed here. Further, if some policies are introduced at the same time, they are likely to target different subpopulations, thus the same parameter for different countries would have a different meaning (put otherwise, it would be valid for different subpopulations that are different for reasons beyond the nationality); if the same policy is introduced with different timing, the parameter could be influenced by the time when it is introduced, which might be a critical aspect during the period of this analysis.

Quarter or semester of birth is widely used whenever it is available (Angrist and Krueger (1991)). EU-SILC contains this information for most but not all countries, among which Germany and UK that are the most relevant in terms of GDP: for this reason, in the analysis I refrain from using it. In Table 4 I report the results using the semester of birth as instrument, for countries where it is available. Kitagawa (2015) shows that quarter of birth satisfies both the exogeneity and monotonicity conditions, which therefore are not tested here. This instrument implies smaller opportunity non monetary costs associated with obtaining an additional year of school for individuals who are already enrolled in the next grade when they reach the minimum age for leaving school than for individuals who are not yet enrolled. Since high ability individuals would continue anyway with education, the IV-LATE parameter shares the same interpretation of family background as instrument (though family background should be more binding on the monetary side of the opportunity costs). It is reassuring that the return on education estimated with the two different instruments are indistinguishable for the high educational level. For the low educational level the two estimates are not much different, with the semester of birth returning a slightly higher coefficients: if anything, this suggests that monetary constraints are more binding than non monetary when it comes to a cost-benefit analysis.

This discussion makes clear that even an appropriate instrument does not solve the limits of the IV-LATE when the interest is in a cross country comparisons.