Document de travail du LEM / Discussion paper LEM
2018- 10

The intergenerational transmission of education. A meta-regression analysis.

Nicolas FLEURY

Center of Studies - Alpha Group, Paris, France and University of Lille / nicolas.fleury@univ-lille1.fr

Fabrice GILLES

LEM UMR 9221 / fabrice.gilles@univ-lille.fr

http://lem.cnrs.fr/IMG/pdf/dp2018-10.pdf

http://lem.cnrs.fr/

The intergenerational transmission of education. A meta-regression analysis. [a]

Nicolas Fleury[b] and Fabrice Gilles[c]

(First version: June 2015; This version: March 2017)

Abstract

In this article, we conduct a meta-regression analysis to evaluate the extent of the impact of parental education on the education of their children. Starting in the beginning of the 2000s, a growing body of literature has focused on the causal impact of parents' education on that of their children, and has provided a large range of values of the education transmission coefficient. We review the empirical literature and propose a multivariate meta-regression analysis to estimate the true effect of parental education. Our database is composed of a large set of both published and unpublished papers written in the period 2002-2014. The data allow us to econometrically evaluate the effect of parents' education on their children's education. The articles considered differ in the data sources, explanatory variables, econometric strategy applied, and the type of publication. In spite of the large heterogeneity of studies and evidence for publication bias, we find a significant transmission of education from parents to their children: this effect is estimated to around 0.18.

Key-words: education, intergenerational transmission, meta-regression analysis.
JEL Classification: C83, J13, J24.

1. Introduction: evaluating the causal impact of parents' education on that of their children

In this article, we evaluate the extent of the effect of parents' education on that of their children by using a quantitative and innovative method, the meta-regression analysis, to review and analyze the empirical estimates of the existing works that deal with this matter.

[a] We thank participants in the *XXIII Meeting of the Economics of Education Association* (Valencia, Spain; July, 2014), the *Meta-Analysis of Economic Research* Network conference (Athens, Greece; September, 2014), the *Direction de l'évaluation, de la prospective et de la performance du Ministère de l'Education nationale* seminar (Paris, French Ministry of Education; January, 2015), the 'SIUTE' seminar (Lille 1 University, France; March, 2015), the *International Workshop on Applied Economics of Education* (Catanzaro, Italy; June 2016), and the *European Association of Labour Economists* (Ghent, Belgium; September 2016) for helpful comments on previous versions of this article. We are also grateful to Erik Plug, Kjell Salvanes, and Tom Stanley and two anonymous referees.

[b] Center of Studies - Alpha Group, Paris, France and University of Lille, Villeneuve d'Ascq, France. E-mail: nicolas.fleury@univ-lille1.fr.
[c] LEM-CNRS (UMR 9221)-University of Lille 1, Villeneuve d'Ascq, France and TEPP. E-mail: fabrice.gilles@univ-lille1.fr.

Since the 1980s, a large strand of empirical work has tried to evaluate the causal impact of parents' education on that of their children. Empirical studies show that (raw) intergenerational *correlations* related to education amount to about 0.4 for Western Europe, 0.46 for the United States, and 0.6 for South America (Black and Devereux, 2011) while *causal* estimates of the effect of parents' education on children's schooling usually exhibit smaller values (Holmlund *et al.*, 2011). Investigating the underlying mechanisms that explain these correlations leads to a "nature *vs.* nurture" debate over the individual accumulation of human capital, and the respective impact of genetics and the child's environment (including parental background). The literature suggests that the search for the causal effect of parental education on children's education corresponds to an investigation into the "nurture effect" (Lochner, 2008; Holmlund *et al.*, 2011). Typically, the coefficient of intergenerational transmission of education (*i.e.* from parents to children) represents this causal effect.

Three main estimation strategies are implemented in the more recent literature to evaluate the "causal effect" of parental schooling (Björklund and Salvanes, 2011; Black and Devereux, 2011; Holmlund *et al.*, 2011): the use of sample data for twins, and for adoptees, and application of instrumental variables (IV) strategies uses normally "representative" samples of the targeted population. However, debate remains open on the size of the causal effect of parental schooling. There is a large uncertainty on the real value of the coefficient of intergenerational transmission of education: a large range of values exists for this coefficient, notably depending of the method used for the estimates. The meta-regression analysis suggests an approach that could be considered as a quantitative survey of the literature (Stanley and Jarrell, 1989; Stanley, 2005; Stanley *et al.*, 2013). Indeed, empirical studies that deal with causal effects of parental education on that of their children are characterized by three main features: *(i)* a large heterogeneity in the datasets used in the literature on the effect of parental schooling; *(ii)* a large variation in the covariates included in the analysis or the econometric methods in us; *(iii)* a potential publication bias in this field of literature.

In that perspective, the meta-analysis represents in particular a framework "*to review more objectively the literature and explains its disparities"* [d] (Stanley and Jarrell, 1989, p.168). Consequently, the present study proposes the first meta-regression analysis (MRA) to a large set of empirical studies (2002-2014) that evaluate the transmission of human capital from parents to their children. This research differs from the previous reviews of the literature like those of Björklund and Salvanes (2011), Black and Devereux (2011) or Holmlund *et al.* (2011) under narrative form[e], because it takes the form of a quantitative survey of the literature based on the existing empirical works. It also brings together both published and *unpublished* papers as long as it follows some strict rules for the inclusion of articles in the set of works considered for the analysis (Stanley *et al.*, 2013). In this paper, the aim is to get a value for the "true" effect of parents' education on that of their children taking account of the heterogeneity of empirical studies and of potential publication bias (Stanley and Jarrell, 1989; Stanley, 2005). The existing empirical studies are characterized by several features including the population group considered, the explanatory variables employed, the econometric strategy, the data sources, and the type of publication. These features may explain why the results of these studies differ (Stanley and Jarrell, 1989; Stanley, 2001). On the other hand, as Begg and Berlin (1988) point out in the case of medical studies, positive results are more likely to be published than studies showing negative results. More generally, published results may overstate or understate the true effect (Stanley and Jarrell, 1989; Ashenfelter *et al.*, 1999, Doucouliagos and Stanley, 2009, Havranek and Irsova, 2011, Huang *et al.*, 2009).

[d] Indeed, the "narrative literature reviews" could *"introduce a substantial bias by omitting portions of the literature, usually on alleged methodological grounds"* (Stanley, 2001, p. 144).

[e] The notable exception is the paper of Holmlund *et al.* (2011) who also propose an original quantitative analysis by implementing the three main identification strategies on a same Swedish dataset.

We contribute to the literature on three levels. First, we show that previous empirical studies have given rise to a large range of values of the education transmission coefficient due to differences in the population studied, the explanatory variables included, the econometric strategy, data sources, and characteristics of the publications. Second, we test for publication bias in the literature on the causal impact of parental schooling on children's schooling. Third, we provide evidence of a genuine empirical effect of parental schooling on children's schooling, net of potential publication bias and heterogeneity of the studies.
The paper is organized as follows. Section 2 presents the motivations of the paper. Section 3 displays the meta-analysis regression dataset and shows heterogeneity in considered empirical studies. Section 4 presents the econometric strategy considering the multivariate meta-regression analysis framework, and provides new evidence for the causal effect of parental education on children's one. Section 5 concludes.

2. Evaluating the causal impact of parents' education on that of their children

In the empirical literature dealing with the estimation of individual human capital accumulation (see e.g. Mulligan, 1997), a child's education attainment is explained by a large set of individual, familial and other environmental variables. The following basic equation can be estimated:

$$EDU_c = \alpha + \beta EDU_p + \gamma X + \varepsilon \qquad (1)$$

where EDU_c is the child's education attainment and EDU_p is the parents' education attainment; X is a set of control variables for individual (e.g. gender or date of birth) or familial (e.g. parents' income, rank among siblings, grandparents' education) features, or geographical variables; and ε is the standard residual.

β is the coefficient of intergenerational transmission of education (or ETC, education transmission coefficient in this article). It refers to the transmission of parents' schooling to their child. In this study, we focus on this coefficient which corresponds to the causal effect of parents' education on that of their children in many existing studies.

Estimating equation (1) to get the causal effects raises at least two main questions.
On the one hand, one relevant question related to the estimation of equation (1) is the following: Should both parents' education be included on the right-hand side of the equation? There is a relative consensus in the literature on the answer whether the spouse's schooling should be included or not, which depends on the goal of the study: what is the question raised? (Björklund and Salvanes, 2011; Holmlund *et al.*, 2011). For instance, do we want to evaluate the intergenerational transmission of education of fathers on that of their children, controlling for assortative mating? If only one parental education is included in the equation, there might exist some difficulty in interpreting the related ETC because of the "assortative mating" phenomenon. It refers to the inclination for one spouse to mate with somebody which has similar socio-economic status (education, income, for instance). This feature may indeed entail large collinearity between parents schooling (Holmlund *et al.*, 2011).

On the other hand, another relevant question is: what kind of identification strategy should be used? As mentioned for instance in Björklund and Salvanes (2011) or in Holmlund *et al.* (2011), three main methods are implemented to estimate the causal effect of parental education on their child's education: the use of an instrumental variables strategy, the use of twins data and the use of adoptees data. It is mainly due to the fact that OLS may in most cases entail biased econometric estimates, and this even in spite of a (potentially) large set of control variables.

The first identification strategy to estimate the coefficient β uses instrumental variables (IV). The related studies generally exploit the existence of natural experiments provided by educational reforms in compulsory schooling that create exogenous variation in parental education. It allows estimating the causal effect of parental education on children's education. The blueprint of these works is the influential study by Black *et al.* (2005) that uses educational reforms from 1960 to 1972 in Norway, extending the "period" of compulsory schooling from the seventh grade to the ninth grade. The authors found strong OLS relationships, but no significant impact (at the 5% level) of parental education considering IV, except for the mother-child relationship of the less educated parents. IV represents an interesting strategy, and with the ability of being replicated for a large number of countries that experienced similar extension of the period of compulsory schooling.[f]
However, this method presents some potential flaws. One major failure is the risk of choosing "bad" instruments: the instruments must be exogenous, and also much correlated with the treatment variable (Wooldridge, 2002). More specifically, most of the literature considers compulsory schooling laws for the secondary school. As mentioned in Black *et al.* (2005), *"It is plausible that a policy change that increased enrollment in higher education would have been transmitted more successfully across generations"* (p. 447): IV estimates could refer to local average treatment effects ("LATE") and may not apply to all kinds of enlarged education policies. In addition, "*the asymptotic variance of the IV estimator is always larger, and sometimes much larger, than the asymptotic variance of the OLS estimator"* (Wooldrige, 2002, p. 467). This last flaw may be particularly problematic for our research question as non-significance would reflect no causal impact of parental education on children education.

The second main identification strategy uses data on adoptees. In this approach, the idea is to capture the effect of parental education on the schooling of children who are not their biological children. Hence, this strategy would allow estimating causal impact of parental schooling net of genetic transmission. Studies in this field use OLS on the sample of adoptees and typically find smaller (and not always significant) values for the ETC of adoptees than for biological children. For instance, Plug (2004) or Sacerdote (2007) use such a strategy. Plug (2004) uses US data (the Wisconsin Longitudinal survey) and finds that heritable abilities ("nature") play an important role in intergenerational transmission, with no causal effect (nurture) of parental education on adoptees. The study by Sacerdote (2007) mobilizes a Korean American dataset and finds a far smaller – but significant – transmission of education coefficient from the considered parents to her/his adopted child relatively to the biological child.
More generally, despite the sophisticated nature of data, there seems to exist some serious potential flaws in that strategy, often mentioned by the authors of the studies. First, the strategy relies on the assumption of randomly assigned parents: the adoptive parents should not be a strongly selected group of parents relatively to rest of the population (Björklund and

[f] Otherwise, rather than considering compulsory schooling, alternative instrumental variables are sometimes used, like the birth order of parents in Havari and Savegnago (2013).

Salvanes, 2011). Second, the size of the considered samples of adoptees is often small: such datasets contains information about a few hundred of individuals and sometimes even less. Third, the age of the adoption matters on educational outcomes. Indeed, especially for adopted children who are not born in the country where they are raised (but not only), there is some acculturation at work (Rumbaut, 2004), depending on the age of arrival (here adoption) that may largely influence education.

The last main identification strategy relies on the use of data on twins. This approach also aims at "controlling" for the "nature" effect to identify differences in schooling of twin parents, by removing the genetic heritable transmission. Studies using this approach typically present OLS estimations and difference between twins (WITHIN estimator). For instance, Behrman and Rosenzweig (2002) use US data from the State of Minnesota and find significant OLS estimates of ETC for twin fathers and twin mothers. They find significant impact of father's education but not of the mother's by using the WITHIN estimator. Holmlund, Lindahl and Plug (2008) use data from Denmark and find similar results, except for the case of assortative mating.
This identification strategy also suffers from important potential flaws. First, non-randomness in the educational choices of the twin is a criterion that has to be met in order to have adequate identification (Björklund and Salvanes, 2011). Second, it seems more relevant to use data on monozygotic twins that dizygotic twins: in the latter case, the twins share different genetic codes which may introduce additional bias in the analysis (Björklund and Salvanes, 2011). Finally, as mentioned in the second approach, the rather small size of the datasets and the possibility of selection in these samples also represent serious potential flaws that may limit the generalization of the results elaborated with such data.

Overall, a large range of values for the estimates of the ETC exists in the literature. This feature is confirmed by studies that implement the three different identification strategies on a same dataset (Holmlund *et al.*, 2011 on Swedish data and Hægeland *et al.*, 2010 on Norwegian data). One solid conclusion from the literature using the three main estimation strategies exposed present values for the ETC that are lower than the coefficients found with (simple) OLS. Meanwhile, there is no clear further conclusion in the results. The estimation value of the ETC varies notably with the econometric strategy, with the data (country, for instance), or with the parent (father or mother?)[g] considered for the intergenerational transmission. All these different features provide support for the implementation of meta-regression analysis that would propose an estimation of the "true effect" of parental schooling on children's education.[h]

3. Meta-regression analysis: a dataset of education transmission coefficient estimates

In this section, we discuss the empirical framework and present the data set on which our MRA is based.[i]

[g] There is also variation in the estimated ETC depending on if there is assortative mating in the estimation performed.
[h] An additional and important argument for implementing of a meta-regression analysis is to control for possible "publication bias" (Sutton *et al.*, 2000; Stanley, 2005), that has been proven as an empirical feature for a number of research questions (Stanley, 2005; Stanley *et al.*, 2008; Ashenfelter *et al.*, 1999, Doucouliagos and Stanley, 2009; Havranek and Hirsova, 2011).
[i] See for instance Stanley *et al.* (2013) for guidelines on this task.

3.1. Studies included in the MRA dataset

All empirical works that estimate an intergenerational coefficient of transmission of education are candidates for inclusion in the meta-regression analysis. To collect the set of studies to be included, we consider works that aim at evaluating the causal impact of parental education on children's education. Thus, we consider papers that adopt one of the three identification strategies aforementioned.[j] Moreover, within a given paper among those we include in our file drawer, we won't consider some results as soon as they are used as a comparison basis. This is mainly the case of regressions that use OLS to get estimates of the beta coefficient: OLS that are used in IV studies, or that applied to own birth child in studies on adoptees or to twins samples.

Besides, to be included the studies should report an explicit value of the effect of parental education on children's education. Parents' and children's education levels should be expressed in years of schooling. Completed years of education are the basis for a large body of empirical work on the individual transmission of education.[k] The coefficient of intergenerational transmission should not correspond to an elasticity: most of this literature does not express years of education in logarithms. Finally, studies which use degree level as the education variable cannot be included because the econometric results in these ordered logit/probit models are not directly comparable with those obtained in linear models (years of schooling).

We performed several searches of scholarly databases and other internet searches between December 2013 and February 2014, using a large set of keywords closely related to the topic of impact of parental education in the human capital literature.[l] First, we searched EconLit databases (Cairn, JSTOR, Science Direct, Springer Link) for published academic papers. Second, we extended the search to specialized research institution websites (IZA, NBER, SSRN) for working papers or research reports on labor/education economics. Third, we searched numerous Google web page results to identify work in progress and other non-published research. Fourth, we tried to ensure that no relevant work was overlooked by searching the references in the selected papers. Where different versions of the same paper exist, we consider the published or most recent version. We consider only papers with cross-sectional data, *i.e.* individual observations covering a fairly large range of birth cohorts, such as samples representative of a population or specific samples (e.g. twins or adopted children) but not one or a small number of cohorts. The final dataset was checked for coherence and for possible errors in the coding of the different variables.

[j] Other ways were recently used to identify the causal effect of parental schooling on that of their children. It is for instance the case of de Haan (2011), who considers bounding the causal effect of parental education on that of their children, by using an analysis based on Manski and Pepper (2000). However, we won't consider such studies because only a limited number of results are available by now. Moreover, some articles also apply IV to twins sample (such as Behrman and Taubman (1985) or some estimates provided in Bingley, Christensen and Jensen (2009); we also exclude these estimates from our file drawer because they refer to strategies that are also rarely used.

[k] Alternative measures of education include highest diploma level achieved by the individual; this will be investigated in future research.

[l] These expressions include: *intra-family transmission of education, intergenerational transmission of education, educational intergenerational mobility, intergenerational education/schooling mobility, educational persistence, correlation between parents and child's schooling or education, intergenerational education correlation, intergenerational effects, intergenerational associations/transmissions, causal effect of parent's schooling on child's schooling, intergenerational schooling associations, transmission of human capital/education, causal relationship between parents' and children's education, and accumulation of human capital.*

Our final dataset contains information on 25 articles published or written in the period 2002-2014. This set of 205 estimates of the education transmission coefficient corresponds to effect size (Stanley and Jarrell, 1989). A given effect size corresponds to an estimate of the intergenerational transmission of education from parents to their children in the estimation of equation (1).

Table 1 presents the articles in our dataset. We find an average of 18 estimated values for the effect size for each study. Finally, the average coefficient in the empirical studies is 0.14.

Table 1. Studies included in the meta-regression analysis.

Author(s)	Nb. of effect sizes in study	Average effect size
Antonovics and Goldberger (2005)	2	0.254
Amin, Lundborg and Rooth (2011)	28	0.079
Aguero and Ramachandran (2010)	2	0.078
Bingley, Christensen and Jensen (2009)	8	0.057
Black, Devereux and Salvanes (2005)	6	0.054
Bjorklund, Janti and Solon (2007)	12	0.068
Björklund, Lindhal and Plug (2004)	12	0.099
Björklund, Lindahl and Plug (2006)	8	0.093
Behrman and Rosenzweig (2002)	14	0.035
De Haan (2008)	8	0.165
Hoffman (2013)	6	0.030
Holmlund, Lindhal and Plug (2011)	12	0.071
Havari and Savegnago (2013)	6	0.455
Kallioniemi (2014)	4	0.044
Lindhal, Palme, Massih and Sjögren (2013)	2	0.315
Meng and Zhao (2013)	15	0.330
Plug (2004)	6	0.196
Pronzato (2012)	4	0.129
Plug and Vijverberg (2005)	6	0.158
Sacerdote (2000)	2	0.192
Sacerdote (2004)	3	0.056
Sacerdote (2007)	2	0.093
Schultz (2004)	8	0.343
Stella (2013)	5	0.410
Tsou, Liu and Hammitt (2012)	24	0.075
Sample average	**8.1**	**0.135**

Sources: Authors'compilation. See Appendix for full references.

3.2. Descriptive statistics: heterogeneity in estimated values of ETC

The estimated coefficient of parental transmission of education is provided for a specific estimation within a given study. In particular, this coefficient is related to a specific survey, a given population of children or parents, a given set of control variables included when estimating the effect size for a given econometric estimator, and/or a particular type of publication. Therefore, a given estimated value of the education transmission coefficient might be linked to all of those specific features.

In what follows, we propose a definition and coding of the moderator variables, *i.e.* variables that describes empirical studies and may be relevant for explaining why effect size β_j might differ across empirical studies.

Those meta-independent variables belong to one of seven groups of variables:

- *Data type*: contains general information on data sources (country surveyed).
- *Children*: provides information on characteristics (twins or adopted children; ethnic origin; boy or girl, etc.) considered to get estimate of β .
- *Parents*: provides general information on the parent considered (*e.g.* mother or father) to get estimate of β .

- *Socioeconomic control variables*: to indicate whether or not some control variable are included in the econometric specification to arrive at a given estimate of β. The control variables include children's characteristics (age, gender) or their family's characteristics (household income, number of siblings, etc.). Geographical indicators are also taken into account.
- *Estimator*: refers to the econometric methodology used in the research.
- *Publication characteristics:* characterizing the type of publication: academic journal, working paper, book chapter, conference proceedings.

Appendix Table A1 provides definitions and sample statistics (means and standard deviations) for those variables for the full set of publications in our meta-database.

Table 2 reports the mean difference between the effect sizes for the target group and the effect sizes for the remaining (reference) group of estimated effect sizes for every moderator variable that corresponds to a dummy variable.

Table 2. Difference in the mean effect size by characteristics of the study.
Population: full sample (all publications).

Variable name	Difference[a]	Significance[b]
Data type:		
Africa	0.163 (0.057)	0.018**
America	-0.019 (0.027)	0.492
Asia	0.013 (0.024)	0.590
Europe	-0.027 (0.020)	0.176
Children:		
Boy	0.003 (0.024)	0.910
Girl	-0.059 (0.028)	0.039**
All gender	0.037 (0.021)	0.075*
Parents:		
Mother	-0.082 (0.019)	<0.001***
Father	-0.021 (0.019)	0.266
Both parents	0.216 (0.029)	<0.001***
Socioeconomic control variables:		
Gender	0.043 (0.021)	0.037**
Age/Birth	-0.021 (0.037)	0.573
Number of siblings	0.069 (0.038)	0.075*
Rank among siblings	-0.027 (0.025)	0.288
Assortative	-0.070 (0.019)	<0.001***
Birth parents	-0.102 (0.016)	<0.001***
Professional status	0.318 (0.016)	0.019**
Income	0.036 (0.038)	0.352
Local	0.062 (0.022)	<0.001***
No covariates	0.020 (0.135)	0.889
Estimator:		
OLS	-0.064 (0.018)	<0.001***
IV	0.172 (0.026)	<0.001***
Within	-0.085 (0.020)	<0.001***
Publication characteristics:		
Academic	-0.015 (0.020)	0.450

Notes: [a] Difference refers to the mean difference between the effect sizes for the target group and the effect sizes for the remaining (reference) group.. [b] P-value (probability to reject the alternative hypothesis) for the statistical significance of the group difference. *** (resp. ** and *) stands for significance at a 1% (resp. 5% or 10%) level.

For instance, relatively to any other group of countries, the effect size on average is greater for estimates that rely on Africa related data. Also, relative to other children, effect size is smaller for girls. The same holds when considering ETC evaluated for mothers. Besides, effect size tends to be larger among estimations that include (in the specification of the estimated equation) variables related to the gender of the individual, or the professional status of her / his parents. On the contrary, taking account in the list of the explanatory variables of factors related to the birth of parents or the education level of the spouse (assortative mating) of the considered parent seem to be negatively correlated with the estimated ETC. Effect sizes are also lower on average for estimates based on the WITHIN estimator (thus while considering estimates on twins), or on the OLS estimator (thus using data on adoptees); the contrary holds for IV estimates, in comparison with OLS or WITHIN estimates. Hence, effect size varies with the characteristics of studies in most cases.[m]

The next step in our analysis is multivariate meta-regression to take account of this heterogeneity.

4. Exploring heterogeneity from education transmission coefficient: a multivariate meta-regression analysis

In Section 3, we showed that the set of studies considered displays a large range of values for the effect size (see Table 1). We also show that all those studies are characterized by several specific features that may explain why estimations of the ETC might differ (Table 2).

In this section, we provide some evidence of potential publication bias, and a genuine empirical effect on the education transmission coefficient (taking no account of the heterogeneity of studies). In particular, as it is usual in Multivariate Regression Analysis (Doucouliagos and Jarrell, 2009; Huang *et al.*, 2009; Stanley, 2005), we disentangle publication bias and the genuine empirical effect through the application of funnel asymmetry testing, while taking account of the heterogeneity of studies (FAT-MRA).

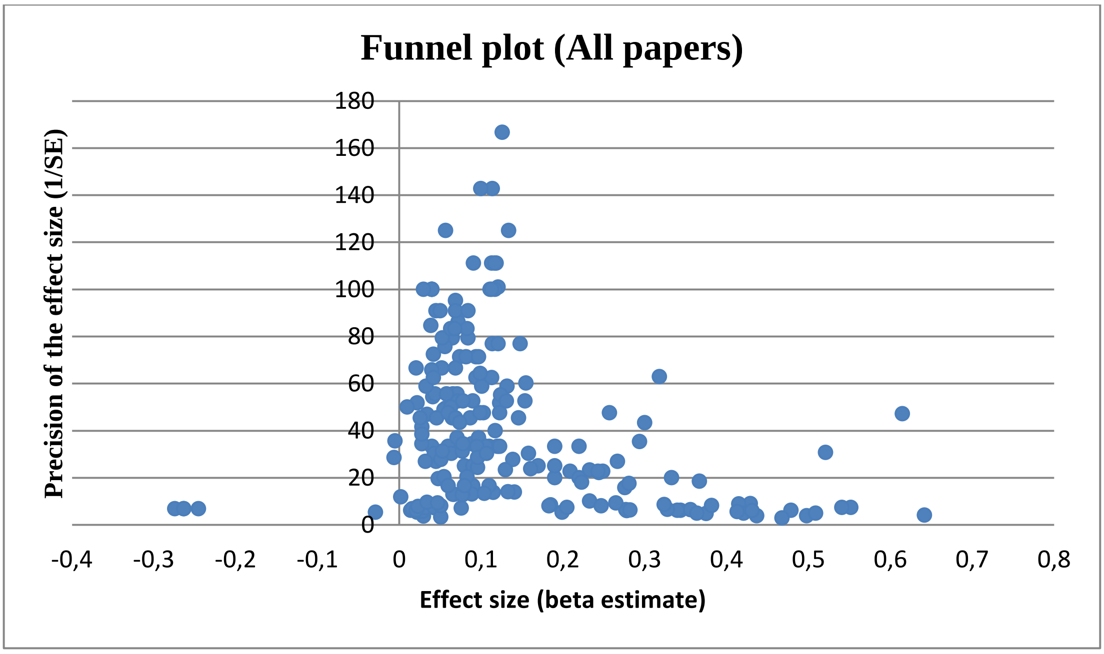
4.1 Publication bias: funnel asymmetry

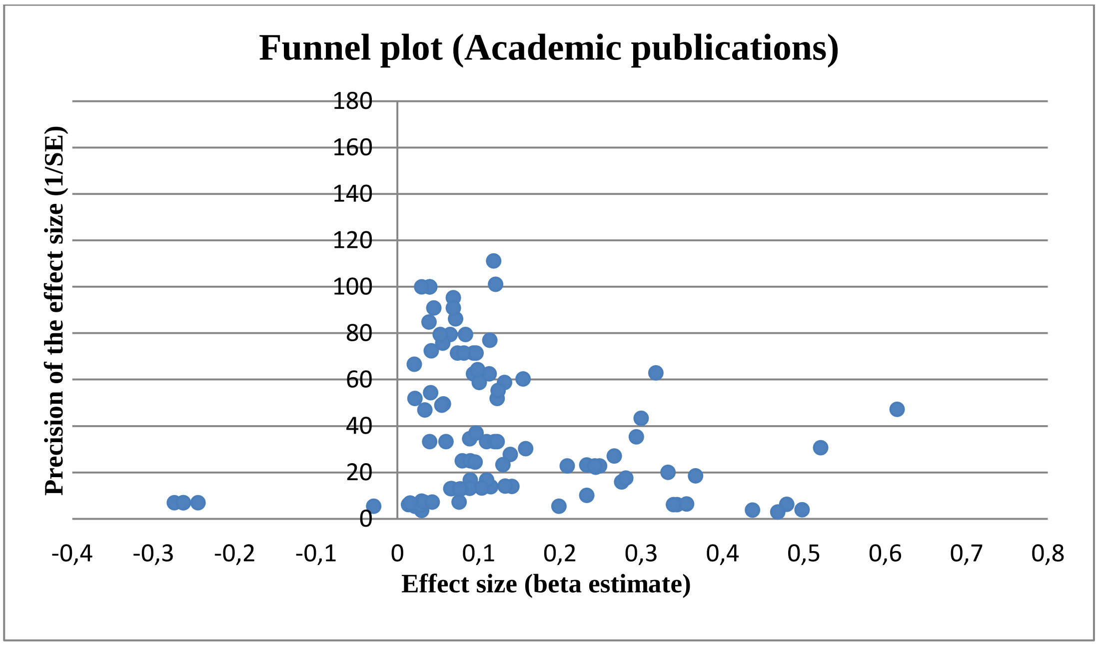
The literature that deals with MRA to estimate a genuine effect often distinguishes between a true effect and publication bias. Indeed, Begg and Berlin (1988) in the case of medical studies show that papers that provide positive results (*i.e.* indicating a positive effect of the 'treatment') are more likely to be published. More generally, and particularly in economics, published results can overstate or understate the true effect (Stanley and Jarrell, 1989; Huang *et al.*, 2009) such that the estimated effects of parental schooling might be correlated with sampling errors. If these effects are correlated with other variables, then the conclusions about the determinants of children's schooling may be seriously biased. The existence of such bias is due to the natural workings of a scientific process designed to discover important new results (Ashenfelter *et al.*, 1999).
Funnel plot is a first approach to detecting publication bias (Sutton *et al.*, 2000; Stanley, 2005). For all the studies in the MRA dataset, it displays an empirical relationship between the estimated beta coefficient and its precision (usually the inverse of the standard error

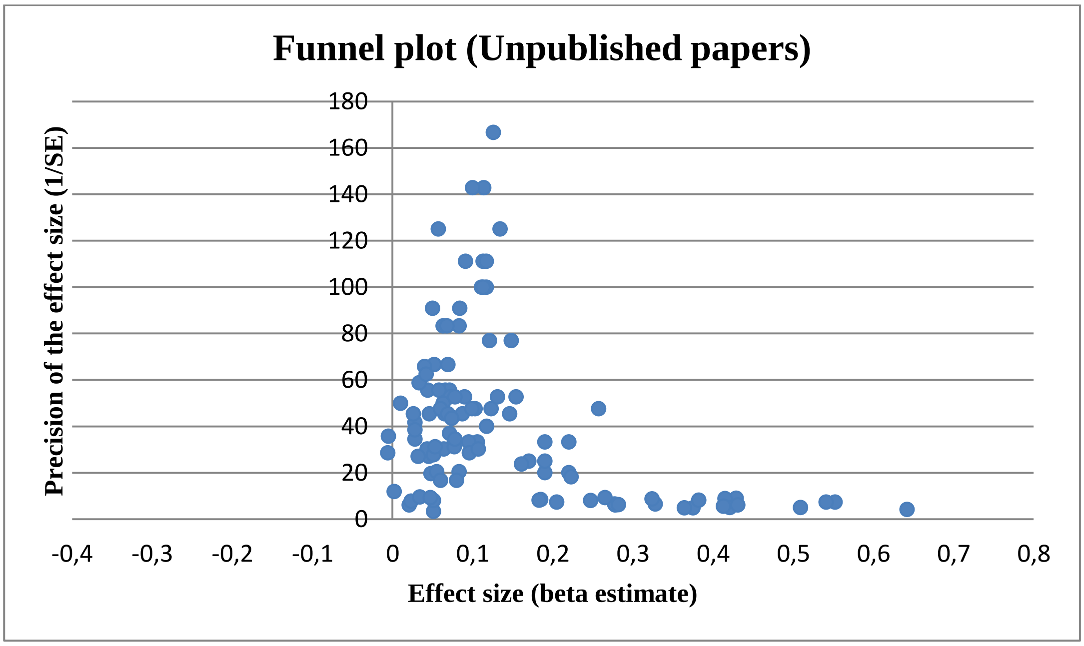
[m] Considering separately the sample of academic publications or of unpublished papers also shows a large heterogeneity in the estimated effect. Moreover, both kinds of sample are different but share a several features for effect size. Corresponding Tables are available on request.

estimate). Sutton *et al.* (2000) refer to an overweighted plot on one side or another around the 'true effect' of parental education could be a sign of the existence of publication selection. Thus, we first perform funnel plots on the whole sample, and then we consider two article type sub-samples: academic publications, and unpublished papers (including working papers). Figure 1 shows that there may be some publication bias: considering the whole sample leads to an overweighting on the right side, even if it is not so clear-cut. This appears to be due to the effect of "unpublished papers", for which there are also sometimes large values for effect size, but also for which precision is often greater than for academic papers, even for estimated effect of small size.

Figure 1. Funnel plots for the intergenerational transmission of education, different sub-samples of observations.







4.2 Multivariate MRA, publication bias, and genuine effect

Funnel plots are graphs. We can perform a formal test for graph asymmetry (Stanley, 2005). The starting point for FAT is the relationship between the reported coefficient of parental transmission of education and its standard error (Egger *et al.*, 1997):

$$b_j = \beta_1 + \beta_0 SE_j + u_j \quad (2)$$

where b_j denotes the estimated coefficient of transmission of education from parents to their children. This coefficient is reported in the j^{th} study in our final dataset (*j= 1, 2, ... N*). SE_j is the standard error of b_j, and u_j is a random residual. If there is no publication bias, the estimated effects should vary randomly around the genuine value of the coefficient β_1.
The FAT consists of a *t*-test performed on the intercept (β_0). If β_0 is different from zero, there is evidence of funnel symmetry, and therefore publication bias.

So far, to disentangle empirical effect and publication bias, we also have to take account of heterogeneity of studies. Indeed, this may be of importance because Table 4 in Section 3 shows that effect sizes seem to differ across studies according to several features of articles. The FAT-MRA approach generalizes the FAT analysis. It allows us to estimate the "true" (or "genuine") effect of parental education on children's education, *i.e.* net of heterogeneity of the studies and publication bias (Stanley, 2005). To proceed, we generalize equation (2) and add the moderator variables, *i.e.* dummies that take account for features of the considered studies, divided by the effect size standard error (Stanley *et al.* (2008), for instance):

$$b_j = \beta_1 + \beta_0 SE_j + \sum_{k=1}^{K} \alpha_k Z_{jk} + u_j \quad (3)$$

where Z_{jk} are the *K* moderator variables or meta-independent variables (Stanley, 2001). β_1 represents the "true" value of the transmission coefficient, and β_0 refers to publication bias. Finally, u_j is the meta-regression disturbance term.

Since economic studies use different sample sizes and different econometric models and techniques, u_j are likely to be heteroskedastic. To cope with this problem, we apply OLS to equation (3) where all the terms are divided by SE_j:

$$t_j = \beta_0 + \beta_1 \left(1/SE_j\right) + \sum_{k=1}^{K} \alpha_k \left(Z_{jk}/SE_j\right) + \varepsilon_j \quad (4)$$

Where $t_j = b_j / SE_j$ and $\varepsilon_j = u_j / SE_j$.
Several of the studies in our dataset report more than one and sometimes a large number of estimated values for the beta coefficient. We handle this by clustering the standard errors to avoid the FAT being biased (Sterne *et al.*, 2000; Macaskill *et al.*, 2001). Thus, equation (4) is estimated using OLS and considering clustered standard errors at the study level. Finally, to model publication bias further, we add a dummy to take account for unpublished works (*ie* not peer-reviewed journals).

4.3 Results

Table 3 presents evidence of publication bias and empirical genuine effect after controlling for heterogeneity of empirical studies.

First, the most important scientific issue when considering the transmission of education from parents to their children is the size of the "genuine empirical" effect. Table 4 shows a value for the β_1 coefficient at 0.178. In the MRA approach, as we mentioned it earlier, it is important to consider the wider set of studies that deal with the matter we focus on, and not only the academic publications. Thus, as an estimation of the "true" effect of parental education on that of their children, we consider the coefficient that is related to the full set of papers of our file drawer that is about 0.18. This value does not depend on the characteristics of studies (kind of sample, data type, control variables under consideration, econometric estimator, and so on), as well on the fact that the considered articles are published or not; see Stanley, Doucouliagos and Jarrell (2008).

Table 3. Multivariate Meta-Regression Analysis of the effect of parental education on children's education.

Moderator variable	**Estimates**
Intercept	0.623 (0.610)
Precision (1/Se)	0.178*** (0.059)
Data type:	
Africa	0.010 (0.077)
America	Ref.
Asia	0.044 (0.061)
Europe	0.031 (0.068)
Children:	
Boy	0.060 (0.077)
Girl	0.059 (0.077)
All gender	Ref.
Parents:	
Mother	-0.006 (0.012)
Father	0.011 (0.010)
Both parents	Ref.
Socioeconomic control variables:	
Gender	0.049 (0.071)
Age/Birth	0.052*** (0.013)
Number of siblings	-0.047*** (0.016)
Rank among siblings	0.031 (0.067)
Assortative	-0.011 (0.012)
Birth parents	-0.110 (0.067)
Professional status	0.367*** (0.124)
Income	-0.033 (0.036)
Local	0.011 (0.028)
No covariate	0.014 (0.063)
Estimator:	
OLS	-0.103* (0.054)
IV	Ref.
Within	-0.200** (0.094)
Publication characteristics:	
Academic	Ref.
Unpublished	-1.380** (0.050)
Number of estimates	205
R^2	0.713

Notes: Dependent variable is the t-statistic (effect size related to effect standard error). WLS estimates with clustered standard errors are computed at the study level. Some publication variables (fields of research, number of citations and social science impact factors) are only available for journals (academic publications).

Second, we also examine the possible publication bias. Publication bias is measured by the intercept in Multivariate MRA approach. According to Table 4, estimated intercept is not significant. Thus, it indicates no publication bias. There seems to be no publication bias

within the strand of literature that is related to the estimation of the causal effect of parents' education on that of their children.
Third, the results displayed in Table 4 also provides evidence for the fact that the empirical effect size (*i.e.* each value of the effect found in the existing studies) is partly explained by heterogeneity among studies that analyze the impact of parents' education on children's education. For instance, *ceteris paribus*, effect size is of the same order, either the education transmission coefficient is related to both parents, or to one of them only. Among the socioeconomic control variables, effect size is smaller in studies that take account for the number of siblings. On the contrary, effect size is greater in articles where estimations include as explanatory variables information related to the age of the individual, or related to the professional status of her/ his parents. Moreover, *ceteris paribus*, effect size is lower among estimates that were obtained using OLS (thus based on a sample of adoptees), or WITHIN (thus based on a sample of twins) estimators, rather than considering the IV estimator.

Overall, we still find a significant and positive genuine empirical effect of parental education children's education, irrespective of publication selection and heterogeneity of studies.

Finally, it should be noted that the R-squared of both regressions is high (larger than 0.71): this indicates that our models explain the main part of the heterogeneity in the estimated coefficient of transmission of education attainment.

4.4 Discussing the main results

The FAT-MRA regressions give evidence for a causal effect of parental schooling on their children's education. The transmission education coefficient is estimated to around 0.18, which is of significant magnitude and in the MRA literature corresponds to the "true" or "genuine" empirical effect of the interest variable (*e.g.* Stanley, 2005). However, the FAT-MRA on the three considered samples shows that a large number of moderators are significantly related to the estimated coefficient of parental transmission of education. Hence, the heterogeneity of studies explains a large part of the variation in the coefficient of parental transmission of education in related empirical studies. Moreover, there is evidence for publication bias. Thus our results show it was important to consider meta-regression analysis to provide new evidence on the causal effect of parental education on that of their children.

5. Conclusion

In this article, we provide new evidence for the causal effect of parental education on children's education. Indeed, since the beginning of the 2000s there is a growing strand of literature that aim at evaluating this effect. Corresponding articles are characterized by a large range in estimated values of the education transmission coefficient. Thus, we built a large set of empirical studies that aim at evaluating the causal effect of parental education on that of their children. We then conducted a multivariate meta-regression analysis to estimate the causal effect of parents' schooling on that of their children, irrespective of the heterogeneity among the studies considered, or any potential publication bias.
Multivariate meta-regression analysis showed that the heterogeneity of the studies explains a part of the variation in the estimated coefficient of parental transmission of education in considered empirical studies. Finally, we found evidence of a significant and positive causal

effect of parental schooling, net of potential publication bias and of the heterogeneity of the studies. This amounts to 0.18. This is an important result for public policy actions: raising parental levels of education clearly benefit to offspring's education, through direct effect.

In our study, we focus on years of schooling. However, there are alternative measures of education. They include highest level of diploma achieved by the individual. Thus, if we can find a sufficiently large sample of studies in this particular strand of literature, it would be interesting to analyze the impact of parental education on the probability for her/his child of achieving a low or a high level of diploma. This will be investigated in future research.

References

Agüero J. M. and **Ramachandran M.** (2010), "The Intergenerational Effects of Increasing Parental Schooling: Evidence from Zimbabwe", University of California, mimeo.

Amin V., **Lundborg P**. and **Rooth D.O.** (2011), "Mothers Do Matter: New Evidence on the Effect of Parents' Schooling on Children's Schooling Using Swedish Twin Data", IZA Discussion Papers 5946, Institute for the Study of Labor (IZA).

Antonovics K.L. and **Goldberger A.S.** (2005) "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Comment", *American Economic Review*, 95(5), 1738-1744.

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

Begg C. B. and **Berlin, J. A.** (1988), 'Publication bias: a problem in interpreting medical data', *Journal of the Royal Statistical Society*, (Series A), 151, 419-445.

Behrman J.R. and **Rosenzweig M.R.** (2002), "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?", *The American Economic Review*, 92(1), 323-334.

Behrman J.R. and **Taubman P.** (1985), "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model", *The Review of Economics and Statistics*, 67(1), 144-151.

Bingley P., **Christensen K.** and **Jensen V.M.** (2009), "Parental Schooling and Child Development: Learning from Twin Parents", Social Policy and Welfare Services Working Paper 07:2009, Danish National Centre for Social Research.

Björklund A., Lindahl M. and **Plug E.** (2004), "Intergenerational Effects in Sweden: What Can We Learn from Adoption Data?", IZA Discussion Papers 1194, Institute for the Study of Labor (IZA).

Bjorklund A., **Lindahl M.** and **Plug E.** (2006), "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data", *The Quarterly Journal of Economics*, 121(3), 999-1028

Björklund A., **Jäntti M.** and **Solon G.** (2007), "Nature and Nurture in the Intergenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents", *The B.E. Journal of Economic Analysis and Policy*, 7(2), 1-23.

Black S. and **Devereux P.** (2011), "Recent Developments in Intergenerational Mobility"», in Ashenfelter and D. Card eds, *Handbook of Labor Economics*, ed. 1, 4(5), Elsevier.

Black S.E., **Devereux P.J.** and **Salvanes K.G.** (2005), "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital", *The American Economic Review*, 95(1), 437-449.

Doucouliagos H. and **Stanley T. D.** (2009), "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis," *British Journal of Industrial Relations*, 47(2), 406-428.

Egger M., **Smith G.D.**, **Scheider M.** and **Minder C.** (1997), "Bias in meta-analysis detected by a simple, graphical test", *British Medical Journal*, 316, 629-634.

de Haan M. (2008), "Family background and children's schooling outcomes", Dissertation, University of Amsterdam.

Hægeland T., **Kirkebøen L. J.**, **Kirkebøen O. R.**, and Salvanes K. G. (2010), “Why Children of College Graduates Outperform their Schoolmates: A study of Cousins and Adoptees”, IZA Discussion Paper, 5369.

Havari E. and **Savegnago M.** (2013), “The causal effect of parents’ schooling on children’s schooling in Europe. A new instrumental variable approach”, mimeo, University of Venice, University of Rome.

Havranek T. and **Irsova Z.** (2011), “Estimating vertical spillovers from FDI: Why results vary and what the true effect is”, *Journal of International Economics,* 85, 234-244.

Hoffman A.B. (2013), “Parental Education and Child Human Capital: Evidence from Indonesia”, Master’s Thesis, Lund University, Department of Economic History

Holmlund H., **Lindahl M.** and **Plug E.** (2011), « The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods », *Journal of Economic Literature*, 49(3), 615-51.

Kallioniemi M. (2014), “Adoptee studies and transmission of education”, Master's thesis, Department of Economics, Aalto University School of Business.

Lochner L. (2008), « Intergenerational Transmission », in *The New Palgrave Dictionary of Economics*, Second Edition.

Macaskill P., **Walter, S. D.** and **Irwig L.** (2001), “A comparison of methods to detect publication bias in meta-analysis”, *Statistics in Medicine*, 20, 641-654.

Manski C. and **Pepper J.** (2000), “Monotone Instrumental Variables with an Application to the Returns to Schooling”, *Econometrica*, 68, 997-1010.

Meng X. and **Zhao G.** (2013), “The Intergenerational Effect of the Chinese Cultural Revolution on Education”, mimeo, Research School of Economics, College of Business and Economics, The Australian National University.

Mulligan C. (1997), *Parental Priorities and Economic Inequality,* The University of Chicago Press.

Plug E. (2004), “Estimating the Effect of Mother's Schooling on Children's Schooling Using a Sample of Adoptees”, *The American Economic Review*, 94(1), 358-368.

Plug E. and **Vijverberg W.** (2005), “Does Family Income Matter for Schooling Outcomes? Using Adoptees as a Natural Experiment”, *The Economic Journal*, 115(506), 879-906.

Pronzato C. (2012), “An examination of paternal and maternal intergenerational transmission of schooling”, *Journal of Population Economics*, 25, 591–608.

Rumbaut, R. (2004), “Ages, Life Stages, and Generational cohorts: Decomposing the Immigrant First and Second Generations in the United States”, *International Migration Review*, 38(3), 1160-1205.

Sacerdote B. (2000), "The Nature and Nurture of Economic Outcomes," NBER Working Papers 7949, National Bureau of Economic Research.

Sacerdote B. (2004), “What Happens When We Randomly Assign Children to Families?”, NBER Working Papers 10894, National Bureau of Economic Research.

Sacerdote B. (2007), “How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees”, *The Quarterly Journal of Economics*, 122(1), 119-157.

Schultz T.P. (2004), “Social Value of Research and Technical Skills: Does It Justify Investment in Higher Education for Development?”, *JHEA/RESA*, 2(1), 92–134.

Stanley T.D. (2001), “Wheat from Chaff: Meta-analysis as Quantitative Literature Review”, *Journal of Economic Perspectives*, 15(3), 131-150.

Stanley T.D. (2005), “Beyond publication bias”, *Journal of Economic Surveys*, 19(3), 310-345.

Stanley T.D., **Doucouliagos H.**, **Giles M.**, **Heckemeyer J. H.**, **Johnston R. J.**, **Laroche P.**, **Nelson J. P.**, **Paldam M.**, **Poot J.**, **Pugh G.**, **Rosenberger R. S.** and **Rost K.** (2013), « Meta-analysis of economics research reporting guidelines *», Journal of Economic Surveys*, 27(2), 390-394.

Stanley T.D., **Doucouliagos C.** and **Jarrell S.B.** (2008), « Meta-regression analysis as the socio-economics of economics research », *The Journal of Socio-Economics,* 37(1), 276-292.

Stanley T.D. and **Jarrell S.B.** (1989), « Meta-Regression Analysis: A Quantitative Method of Literature Surveys », *Journal of Economic Surveys*, 3(2), 161-170.

Stella L. (2013), “Intergenerational transmission of human capital in Europe: evidence from SHARE”, *IZA Journal of European Labor Studies*, 2(13).

Sterne J. A. C., **Gavaghan D.** and **Egger, M.** (2000), "Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature", *Journal of Clinical Epidemiology*, 53, 1119-1129.

Sutton A. J., **Duval S.J.**, **Tweedie R.L.**, **Abrams K.R.** and **Jones D. R.** (2000), "Empirical assessment of effect of publication bias on meta-analyses", *British Medical Journal*, 320, 1574-1577.

Tsou M.-W., **Liu J.-T.** and **Hammitt J.K.** (2012), "The intergenerational transmission of education: Evidence from Taiwanese adoption", *Economics Letters*, 115, 134-136.

Wooldridge J. M. (2002), *Introductory Economics*, South-Western College Publishing.

Appendix

Table A1. Summary statistics of the moderator variables in the meta-regression analysis.

Variable name	Variable Description	Mean (Standard Deviation)
Meta-Dependent variable		
T-stat	= Student t-statistic associated to the effect size.	4.16 (0.293)
Meta-Independent variables		
Estimate's accuracy:		
Inverted squared error (ISE)	= Inverted standard error (effect size precision).	40.29 (2.322)
Data type:		
Africa	=1, if the survey deals with a country in Africa.	4.88 (1.51)
America	=1, if the survey deals with a country in America.	20.98 (2.85)
Asia	=1, if the survey deals with a country in Asia.	23.90 (2.99)
Europe	=1, if the survey deals with a country in Europe.	50.24 (3.50)
Children:		
Boy	=1, if the estimated coefficient is only related to boys.	20.98 (2.85)
Girl	=1, if the estimated coefficient is only related to girls.	20.00 (2.80)
All gender	=1, if the estimated coefficient is related to both genders.	59.02 (3.44)
Parents:		
Mother	=1, if the estimated coefficient is related to the mother of the child.	44.39 (3.48)
Father	=1, if the estimated coefficient is related to the father of the child.	41.95 (3.46)
Both parents	=1, if the estimated coefficient is related to both parents.	13.66 (2.40)
Socioeconomic control variables:		
Gender	=1, if the gender is considered as a control variable.	58.54 (2.07)
Age/Birth	=1, if age or birth cohorts are considered as control variables.	81.46 (2.72)
Number of siblings	=1, if the number of siblings is considered as a control variable.	11.71 (2.25)
Rank among siblings	=1, if the rank of the individual among siblings is considered as a control variable.	14.15 (2.44)
Assortative	=1, if assortative mating is controlled for.	49.76 (3.50)
Birth parents	=1, if dummies for parents' year of birth are included as explanatory variables.	40.00 (3.43)
Professional status	=1, if professional status of parents work is included as an explanatory variable.	14.63 (0.84)
Income	=1, if income of parents is included.	14.15 (2.44)
Local	=1, if local dummies are included as control variables.	35.12 (3.34)
No covariates	=1, if no control variables are included.	1.95 (0.97)
Estimator:		
OLS	=1, if OLS estimator is considered.	44.49 (3.48)
IV	=1, if an IV estimator is considered (instrumenting parents' education).	26.34 (3.08)
Within	=1, if a Within estimator is considered (mainly for adoptees data)	29.27 (3.19)
Publication characteristics:		
Academic	=1, if the study is published in an education economics journal.	47.32 (3.50)
Unpublished	=1, if the study is unpublished.	52.68 (3.49)