Université de Cergy-Pontoise
École Doctorale EM2P
(Économie, Management, Mathématiques et Physique)
Laboratoire de recherche THEMA
(THéorie Économique, Modélisation et Applications)

Intergenerational mobility:

An international estimation of extent and determinants of intergenerational transmission of socioeconomic inequalities

THÈSE

pour l'obtention du titre de
Docteur en Sciences Économiques
de l'Université de Cergy-Pontoise
présentée et soutenue publiquement le 26 juin 2017
par

Céline Lecavelier des Etangs-Levallois

JURY:

M. Arnaud LEFRANC, Professeur à l'Université de Cergy-Pontoise, *Directeur*
Mme Christelle DUMAS, Professeur à l'Université de Fribourg, *Rapporteur*
M. Nicolas PISTOLESI, Maître de Conférence à l'Université Toulouse I, *Rapporteur*
M. Markus JÄNTTI, Professeur à l'Université d'Helsinki, *Examinateur*
M. Nicolas FRÉMEAUX, Maître de Conférence à l'Université Paris II, *Examinateur*

Université de Cergy-Pontoise
École Doctorale EM2P
(Économie, Management, Mathématiques et Physique)
Laboratoire de recherche THEMA
(THéorie Économique, Modélisation et Applications)

Mobilité intergénérationnelle :

Une estimation internationale de l'ampleur et des déterminants de la transmission intergénérationnelle des inégalités socio-économiques

THÈSE

pour l'obtention du titre de
Docteur en Sciences Économiques
de l'Université de Cergy-Pontoise
présentée et soutenue publiquement le 26 juin 2017
par

Céline Lecavelier des Etangs-Levallois

JURY:
M. Arnaud LEFRANC, Professeur à l'Université de Cergy-Pontoise, *Directeur*
Mme Christelle DUMAS, Professeur à l'Université de Fribourg, *Rapporteur*
M. Nicolas PISTOLESI, Maître de Conférence à l'Université Toulouse I, *Rapporteur*
M. Markus JÄNTTI, Professeur à l'Université d'Helsinki, *Examinateur*
M. Nicolas FRÉMEAUX, Maître de Conférence à l'Université Paris II, *Examinateur*

Remerciements

Je tiens à remercier toutes les personnes m'ayant aidée pendant l'élaboration de cette thèse et en premier lieu mon directeur de thèse, Arnaud Lefranc, pour son travail de supervision, ses conseils avisés et relectures méticuleuses. Je remercie également les membres de mon jury : Christelle Dumas et Nicolas Pistolesi pour m'avoir fait l'honneur d'être rapporteurs de ma thèse, pour leur temps et l'intérêt porté à ce travail de recherche, Markus Jäntti pour ses remarques et sa participaton scientifique à cette thèse et Nicolas Frémeaux pour avoir accepté aussi de faire partie de mon jury et son intérêt pour mon travail.

Ce travail n'aurait pas été possible sans le soutien du ThEMA, de l'École Doctorale EM2P et du Labex MME-DII. Pour les moyens humains, matériels et financiers dont j'ai pu profiter pour mener à bien mon projet, je remercie leurs directeurs et directrices successifs : Olivier Donni, Mathieu Martin et Olivier Charlot pour le ThEMA, Cristina Terra et Gabriel Desgranges pour l'ED. Je remercie également pour leur travail toutes les secrétaires, passées et présentes, que j'ai eu l'occasion de cotoyer. Je tiens également à exprimer ma gratitude à Michalela Engelmann pour son aide dans l'obtention des données allemandes utilisées dans cette thèse.

Pour les discussions que j'ai pu avoir avec eux et leurs suggestions qui m'ont aidée à améliorer mon travail, je remercie Daniel Schnitzlein, Pierre André, Maëlys de la Rupelle et Thomas Brodaty. Mes collègues doctorants occupent également une place importante dans ces remerciements, tout particulièrement Marie Baguet, pour son amitié et notre collaboration. J'ai une pensée chaleureuse pour ceux qui m'ont accompagnée en cette fin de thèse, Roxane Bricet, Emeline Limon, Mehdi Ayouni, Lionel Lescene, Erwan Moussault, Amélie Adeline, Delphine Drouet, mais aussi pour ceux qui étaient là au début, Charles-Olivier Amédée-Manesme, Romain Legrand, Samia Badji et Ignacio Inoa.

Au terme de ce parcours, je remercie enfin ma famille pour son soutien et ses encouragements. Plus particulirement, merci à ma mère, Nathalie, pour sa présence et sa patience sans faille depuis toujours et tout au long de ce doctorat. Merci à mon père, Hervé, pour son aide et son accompagnement dans le parcours notamment scientifique qui m'a menée jusqu'à cette thèse. Merci à mon frère, Aurélien, pour son exemple, pour nos longues discussions et notre complicité. Merci aussi à mes grands-parents, oncles et tantes, cousins et cousines pour tout ce qu'ils m'ont apporté et sans qui je n'en serais pas là aujourd'hui.

Contents

List of Tables

List of Figures

Introduction

The extent of intergenerational socioeconomic mobility in a society affects the evolution of inequality over time and the way public policies may alter the level of inequality in a given population. The degree to which socioeconomic success is transmitted across generations indeed captures the impact family background can have on children's success in later life: the effect of factors independent from children's choices, talents and efforts on their future success. In other words, it represents to what extent childhood circumstances are reflected in adult life. If family background plays an important role in the success in life, so that individuals born in less advantaged families face worse life prospects than luckier people, one should seek to understand the underlying factors of this transmission of inequality, and public policies might be required to assist mobility. If on the contrary a society is characterized by a high level of socioeconomic mobility, then policy intervention might not be considered necessary. As transmission of inequality – measured by education, occupational status, social class, individual earnings or family income – sheds light on the level of equality of opportunity (Roemer, 1998, 2004, 2012), this subject has long interested sociologists and economists, as reviewed in Björklund and Jäntti (2000), Blanden (2013) and Torche (2015).

In this thesis, we investigate the extent and determinants of the transmission of socioeconomic success from one generation to the next. We first investigate the earnings transmission from fathers to sons in Germany, carefully addressing the question of biases in the estimation. However, the intergenerational elasticity fails at taking account of all factors from the socioeconomic background of an individual affecting future success in life. We then consider sibling correlations as a broader indicator of all family influences, first in France, for different measures of adult attainment: education, occupation and earnings. We also conduct a comparative study of the brother earnings correlation in France and Sweden to assess the impact on the estimation of the lack of information about permanent earnings and the use of predicted measures instead. Finally, we address the question of the mechanisms underlying the transmission of inequality. We thus explore the possibility to use the events of May 1968 in France as a natural experiment to identify and measure the causal link between parental and children's education.

The standard economic theory of income distribution and transmission across generations has been developed by Becker and Tomes (1979, 1986). It has been reassessed notably in Mulligan (1997) and extended more recently in Solon (2004) to explain variations across space and time in intergenerational mobility. It assumes that parents choose to allocate time and money to their own consumption and to the investment in the human and non-human capital of their children. Parents indeed care about and can influence the future success of their children through either their future earnings capacity or direct income transfers. Other factors beyond parental control are also at stake in the determination of the child's earnings. Indeed parents can invest in particular in the health and education of their children, but their choice is determined not only by preferences but also by constraints, such as investment opportunities, and by the rate of returns of these investments. Besides children's earnings are also influenced notably by market luck and inherited abilities, and by social institutions having a role in the degree to which they are transmitted across generations and the impact they have on income (Corak, 2004). The model developed in Solon (2004) shows that intergenerational income transmission "increases with the heritability of income-related traits, the efficacy of human capital investment, and the earnings return to human capital and it decreases with the progressivity of public investment in human capital". This model explains cross-country differences and changes over time, as the mentioned factors affecting intergenerational income mobility can differ across space and time.

The empirical assessment of the degree of intergenerational transmission of inequality is based on this theoretical framework. It has mainly been investigated through the estimation of the empirical association between a given parental measure of socioeconomic success and the children's one. Sociologists commonly evaluate the transmission of the occupational status or class, as reviewed in Treiman and Ganzeboom (1990), Erikson and Goldthorpe (1992, 2002) and Breen (2004), whereas economists are mostly interested in the transmission of individual earnings or family income, as reviewed in Solon (1999, 2002), Björklund and Jäntti (2009) and Black and Devereux (2011). The main advantages of occupational measures of success rely in the facility of data collection (notably as people are more willing to reveal their occupation than their income), and the reliability and stability of the collected information. Occupational status remains relatively stable over the career, contrary to earnings – thus a single observation is more informative and longitudinal data are not required – is better recalled, and adult children can retrospectively report information about their parents. However the occupational class is a highly aggregated measure of success and thus can lead to miss important within class variation. In economic analyses, the log-linear intergenerational income regression commonly estimated provides the intergenerational elasticity (IGE) – or the intergenerational correlation (IGC), if the coefficient is adjusted for differences in income variance across generations – as a measure of intergenerational economic mobility.

Different axes of research have been explored in the estimation of intergenerational mobility. First, even if most of the early studies focused on the United States, other countries have also been considered and several international comparisons of the relative level of intergenerational mobility in different countries have been implemented. They consist either in direct comparative studies, as Björklund and Jäntti

(1997) (Sweden and US), Couch and Dunn (1997, 1999), (Germany, UK and US), Grawe (2004) (Canada, Ecuador, Germany, Malaysia, Nepal, Pakistan, Peru, UK and US), Blanden (2005) (Canada, Germany, UK and US) and Jäntti, Bratsberg, Røed, Raaum, Naylor, Österbacka, Björklund, and Eriksson (2006) (Denmark, Finland, Norway, Sweden, UK and US), or in reviews of the existing literature concerning various countries, as Solon (1999, 2002) (Canada, Finland, Germany, Malaysia, South Africa, Sweden, UK and US), Corak (2006, 2013) (Australia, Canada, Denmark, Finland, France, Germany, Italy, Japan, New Zealand, Norway, Sweden, UK and US) and Björklund and Jäntti (2009) (Australia, Canada, Denmark, Finland, France, Germany, Italy, Norway, Sweden, UK and US). Such comparisons might improve the understanding of the relative importance of the different channels of transmission by assessing possible explanations for differences across countries. They provide a mobility ranking showing intergenerational earnings persistence relatively strong in the United States, low in the Nordic countries, with other western societies in between, such as the United Kingdom, Italy (Mocetti, 2007; Piraino, 2007), France (Lefranc and Trannoy, 2005; Lefranc, 2011), Germany, Japan, New Zealand, Australia or Canada. Grawe (2004) additionally reveal much lower mobility levels in developing nations than in developed countries.

Differences are observed across space, but changes in the degree of intergenerational persistence have also occurred across time and been investigated. They can be associated with evolutions in policy and public spending, and as such might also help to understand the role of different channels of transmission and identify efficient public policies to implement to assist mobility. For the case of the United States, using different sampling and estimation strategies, Levine and Mazumder (2002) found decreasing mobility between fathers and sons, whereas Fertig (2003) found increasing mobility, both observing sons born between the 1940s and the 1970s, and Mayer and Lopoo (2004, 2005) found a decrease of mobility for sons born during the early 1950s and an increase afterwards, for sons born until the early 1960s.

Later work, such as Hertz (2007) and Lee and Solon (2009) find a relatively constant level of mobility for sons born from between the 1950s and the early 1970s, whereas Aaronson and Mazumder (2008) reveal a rise followed by a sharp fall, for sons born between the 1920s and the early 1970s. Thus uncertainty remains about the trends in the intergenerational economic transmission, as no clear consensus has yet emerged. The same conclusion goes for the United Kingdom, where studies either exhibit an increase of mobility, as in Ermisch and Francesconi (2004), or a decrease, as in Blanden, Goodman, Gregg, and Machin (2004) and Nicoletti and Ermisch (2008). In France, Lefranc (2011) observes a fall and a subsequent rise of mobility for sons born between the 1930s and the mid 1970s. Concerning the Scandinavian countries, Pekkala and Lucas (2007) for Finland, Bratberg, Nilsen, and Vaage (2003) for Norway and Björklund, Jäntti, and Lindquist (2009) for Sweden all find evidence of an increasing trend in the intergenerational economic mobility.

Another dimension explored in the analysis of intergenerational economic mobility is the differences across the income distribution. Whereas estimating the intergenerational elasticity only produces a summary measure of transmission from a generation to the next, quantile regressions and transition matrices provide a more detailed picture of mobility. More flexible models of intergenerational transmission than (log-)linear regression are in particular motivated by the fact that, as mentioned, the capacity of parents to invest in their children can be in part determined by credit constraints. Eide and Showalter (1999) use quantile regressions and find that the intergenerational earnings persistence is greater at the bottom of sons' distribution than at the top, in the United States, whereas Corak and Heisz (1999) find lower earnings persistence at the bottom of the distribution than at the top, in Canada, using transition matrices. Couch and Lillard (2004) show that higher-order terms in father's earnings included in the regression equation are typically statistically significant, confirming non-linearities, and find more earnings persistence at the top of fathers' distribution, in Germany and in the United States. For the same countries, Schnitzlein (2015) however finds no evidence for non-linearities along the fathers earnings distribution, while he finds lower persistence at the lowest quartile of the sons distribution, estimating quantile regressions. Grawe (2004) also estimates quantile regressions and find less persistence in upper than in lower quantiles for the United States and Canada, unlike for Germany and the United Kingdom. Exploiting transition matrices, Jäntti et al. (2006) exhibit stronger persistence at the tails of the distributions, particularly at the top for the Nordic countries and the United Kingdom, whereas the persistence tends to be the highest at the bottom of the distribution in the United States. Bratsberg, Røed, Raaum, Naylor, Jäntti, Eriksson, and Österbacka (2007) show that Denmark, Finland and Norway exhibit nonlinearities in earnings mobility – with stronger persistence at the top of the distribution – unlike the United States and the United Kingdom.

Most of the investigation of intergenerational economic mobility estimates the transmission of individual earnings from fathers to sons, excluding women mainly because the labor-force participation of married women is lower than men's, and thus women's earnings are often unreliable indicators of their economic status. However, the analysis has expanded to total family income in order to take non-labor income and individuals without paid employment into account, see for instance Peters (1992). Mothers and/or daughters have also been included, and the role of assortative mating in the transmission process has been considered, notably by Dearden, Machin, and Reed (1997), Chadwick and Solon (2002), Ermisch, Francesconi, and Siedler (2006), Raaum, Bratsberg, Røed, Österbacka, Eriksson, Jäntti, and Naylor (2007) and Hirvonen (2008). Peters (1992) in particular reveals more mobility in earnings than in total income, especially for daughters, in the US. Dearden et al. (1997) find more transmission of earnings (not of education) for daughters than for sons and that, whereas father's education is more important for sons, mother's education and earnings are more important for daughters, in Britain. Chadwick and Solon (2002) exhibit less transmission of family income for daughters than for sons on the contrary, and that assortative mating is an important element in the intergenerational mobility process, in the US. Hirvonen (2008) also finds earnings transmission to be lower for women than for men, replicating the study of Chadwick and Solon (2002) for Sweden. Considering Germany and Britain, Ermisch et al. (2006) also reveal the important role played by assortative mating in the transmission of family income. Raaum et al. (2007) focus on the role of gender and marital status in Denmark, Finland, Norway, the UK and the US. They find in particular that mobility is approximately uniform across countries for married women (contrary to all other groups), when estimates are based on womens earnings, and that married women with children and with husbands from affluent backgrounds present reduced labor supply in the US and the UK, contrary to the Nordic countries.

Thus extensive literature investigated the level of inequality persistence through different approaches, some of them beyond the scope of this thesis. However, empirical difficulties cast some doubt on the reliability or at least the comparability of estimates obtained so far. Indeed, early investigation of the extent of intergenerational earnings mobility in the United States yielded very low estimations of parent-child elasticities around 0.15-0.20 (Behrman and Taubman, 1985; Becker and Tomes, 1986), as they considered a single cross-sectional measure of earnings for each generation, when the interest lies in the transmission of the permanent income (Friedman, 1957), as it is the permanent expectation of income that is assumed to determine consumption and ultimate economic welfare. Thus these analyses led to the conclusion that earnings were not strongly transmitted across generations. Measurement issues in the initial assessments of earnings persistence were then identified and strategies to handle them developed. Even today, the methodological question of the estimation of intergenerational mobility remains essential, in the literature and in this thesis. The main biases arise from transitory fluctuation of current earnings around lifetime earnings (Solon, 1989, 1992; Zimmerman, 1992) and from evolution in the association between current and lifetime earnings over the life-cycle due to heterogeneous age-earnings profiles (Jenkins, 1987; Grawe, 2006; Haider and Solon, 2006). Estimates rose to 0.4 (Solon, 1999) or even 0.5 (Mazumder, 2005), when these biases were taken into account and treated – thanks notably to the progressive availability of larger datasets – revealing much more transmission of inequality than previously estimated.

The first chapter of this thesis analyses intergenerational earnings mobility in Germany, reassessing its extent as uncertainty remains considering the existing literature. Indeed estimates of the intergenerational elasticity or correlation range from 0.11 to 0.46 for this country, in Couch and Dunn (1997, 1999), Lillard (2001), Couch and Lillard (2004), Vogel (2006), Eisenhauer and Pfeiffer (2008) and Schnitzlein (2009, 2015), and are highly sensitive to sample selection and estimation strategies. In particular, attenuation and life-cycle biases can lead to severe misestimation of the economic transmission if not treated. The first bias, the attenuation bias resulting from right-side measurement error in the regression equation of the child's earnings on parental ones, has been discussed in Altonji and Dunn (1991), Solon (1989, 1992) and Zimmerman (1992), and more recently in Björklund and Jäntti (1997), Mazumder (2005) and Nicoletti and Ermisch (2008). This bias results from the use of a single (or two few) parental earnings observation(s) and can be handled either by averaging parental earnings over long periods of time – which is done in this chapter – or by instrumenting them, yet limits of this last strategy are highlighted in Nybom and Stuhler (2011). However life-cycle bias arises not only from the parental, but also from the child's measurement of income, as addressed in Jenkins (1987), Grawe (2006), Haider and Solon (2006), Böhlmark and Lindquist (2006), Brenner (2010) and Nybom and Stuhler (2011), since the association between current and permanent earnings changes over the life-cycle. The extent of life-cycle bias depends on the age at which child's and parent's earnings are observed. A method to correct it is to use earnings observations at a stage in the career when they accurately represent lifetime earnings. Lee and Solon (2009) alternatively control not only for child's age but also for the interaction between child's age and parental income in the estimation of the IGE. Both strategies are implemented in this chapter.

Thus the aim of this chapter is to evaluate the role of father's earnings level on son's one, taking a close look at attenuation and life-cycle biases, using data from the German SocioEconomic Panel (SOEP). We first estimate the association of current and lifetime earnings over the life-cycle to investigate the extent of life-cycle bias from both left and right-side measurement error of earnings, and the best way to treat it, which has never been investigated with data from the SOEP, for Germany. Our results confirm the strong underestimation of the intergenerational transmission arising if sons' earnings are observed in the early stage of the career and that they should rather be observed around the mid-thirties or early forties. Regarding the parental generation, there is a downward bias at any age – partially corrected by the average of parental earnings observations over as long periods of time as possible – however reduced if fathers' earnings are observed around their late forties. We then estimate the intergenerational earnings elasticity around 0.3. We average fathers' earnings over different periods of time to evaluate and reduce attenuation bias, which can lead to an underestimation of the elasticity of 30%. We handle life-cycle bias by controlling son's age span or adding the interaction between son's age and father's earnings in the regression equation. These strategies increase the estimates by 10-20%, correcting some of the life-cycle bias. Our results are in line with the most recent literature about intergenerational economic mobility in Germany and confirm the attention which should be paid to the correct treatment of biases in the estimation of the transmission.

The literature discussed so far has mostly investigated the extent of intergenerational mobility through the association between parental and children's attainments. Indeed one measure of the transmission of economic inequality is whether the earnings of children as adults are closely related to those of their parents. However the impact of the family and community environment on adult success cannot be restricted to a single parental characteristic such as father's income. Another measure of income persistence is whether siblings present a strong resemblance in economic status, showing that family background matters in the determination of adult attainment. If it did not, siblings would not present much more resemblance than any randomly selected unrelated individuals. Sibling correlations thus provide a summary measure of all effects attributed to factors shared by siblings. It captures the overall impact of growing up together, and thus allows a more complete investigation of the role of family background in inequalities. More precisely, the sibling correlation in a particular socioeconomic outcome can be decomposed into the sum of, in the one hand the square of the intergenerational correlation in this outcome and, in the other hand the impact of all other factors shared among siblings and uncorrelated with the specific parental outcome. Therefore, compared to often reported measures of intergenerational elasticity, the sibling correlation in socioeconomic outcomes allows to capture a broader set of family influences.

Sibling correlations have been estimated – in earnings, income or other measures of attainment such as education – in particular in the United States (Corcoran, Gordon, Laren, and Solon, 1990; Solon, Corcoran, Gordon, and Laren, 1991; Levine and Mazumder, 2007; Mazumder, 2008, 2011), in Nordic countries (Björklund, Eriksson, Jäntti, Raaum, and Österbacka, 2002; Raaum, Salvanes, and Sørensen, 2006; Björklund and Jäntti, 2012) and in Germany and Denmark (Schnitzlein, 2014). These studies provide further evidence that family background has a less important impact on adult attainment in Nordic countries than in the United States, with earnings correlations estimated around 0.2 and 0.4 respectively. Besides, sibling correlations shed light on a substantial part of inequality transmission which had not been taken into account by only focusing on the role of parental economic success, revealing much more transmission of inequalities than estimated with intergenerational elasticities. Furthermore, the specific impact in sibling correlations of growing up in the same neighborhood has been assessed, notably by Solon, Page, and Duncan (2000), Page and Solon (2003a,b) for the United States, Raaum et al. (2006) for Norway and Oreopoulos (2003) for Canada. They show that neighborhood does not play a major role and that a large part of neighborhood correlations can be attributed to earnings differential between urban and non-urban areas, combined with the fact that urban children tend to become urban adults.

The second chapter of this thesis examines the contribution of family background to inequality in France by estimating sibling correlations in various measures of socioeconomic success, as no such estimation exists yet for this country. We use data from the French Education-Training-Employment (FQP) survey to investigate similarities between siblings in education, social prestige and earnings, in France. In this country, the intergenerational elasticity has been estimated around 0.5 by Lefranc (2011). Our results indicate a high degree of association in siblings' socioeconomic success. The correlation is around 0.3 and 0.5 respectively for social prestige and years of education. The sibling correlation in annual earnings is around 0.4. All in all, this indicates that estimates of the intergenerational elasticity lead to underestimate the role of family background in children's success in France, as around 30% of the transmission is not accounted for with this measure of intergenerational mobility. We also investigate trends over time in sibling correlations and differences across family types in siblings' characteristics. In particular, we find same-sex and closely spaced siblings to have more in common than brothers with sisters, and siblings with larger age differences, respectively. The size of the family also increases sibling correlations, whereas parental educational and socio-professional levels tend to decrease them.

We then estimate comparable French and Swedish results, both to compare the extent of transmission of inequality in the two countries, and for methodological reasons. For France, due to the lack of extensive enough data, we have to first predict measures of socioeconomic success, in particular earnings, to estimate sibling correlations. However correlations in predicted earnings may not accurately represent actual correlations in permanent earnings, if unobservable characteristics are not associated in the same way than observable ones among siblings. Swedish registers data allow us to assess the impact of using predicted earnings instead of actual permanent earnings and to shed light on the actual extent of earnings transmission in France. Our results confirm the higher level of mobility in Sweden than in France. We estimate the brother correlation in predicted earnings around 0.35 in Sweden, for a correlation in permanent earnings at 0.27. For France, we find a correlation in predicted earnings around 0.54, suggesting a correlation in permanent earnings around 0.4, if the relation between predicted and permanent earnings is assumed to be the same in the two countries.

The previously reviewed literature provides an extensive analysis of the level of inequality persistence due to family background across a number of different dimensions. However it is not only important to assess the degree of intergenerational mobility, but also to understand the mechanisms determining it. There are many channels of transmission and therefore it is essential to distinguish inequalities due to individual preferences, biological or environmental factors. Indeed if individuals choose to follow different paths in life, inequalities resulting from these personal choices are not considered unfair. However if the opportunity for people to achieve their goals is determined by the financial resources of their parents, since parents who achieve economic success can for instance afford the cost of a better education for their children, then public policies may be required to level the field. Nonetheless the difficulty to disentangle all factors raises the question of how to evaluate the causal impact of parents' education on children's one.

To investigate this causal intergenerational transmission of education, there are three different avenues commonly used in the literature, as detailed in Björklund and Salvanes (2010). The first method is based on the study of twins, as in Behrman and Rosenzweig (2002, 2005), Antonovics and Goldberger (2005), Bingley, Christensen, and Jensen (2009), Haegeland, Kirkebøen, Raaum, and Salvanes (2010) and Pronzato (2012). The variation in education among pairs of monozygotic twin parents is used to study the impact on their offspring's education. The purpose in using monozygotic twins is to control for shared environment and genetic material, as it is known to predict abilities. The second strategy relies on samples of adoptees, as in Dearden et al. (1997), Sacerdote (2000, 2007), Björklund, Lindahl, and Plug (2004, 2006), Haegeland et al. (2010) and Holmlund, Lindahl, and Plug (2011). Again the idea is to control for genetics, this time the assumption being the absence of correlation between parent's and child's genetics. For this method to be valid, the family placement should happen in the early stages of life, so that (barely) no investments come from the biological parents, and children need to be randomly allocated to their nonbiological parents to avoid selection bias. The third approach to quantitatively assess the impact of parental schooling on children's one is to exploit exogenous sources of variation in parental education deriving from natural experiments. Minimum school leaving age (MSLA) reforms are widely used, in particular by Chevalier (2004), Black, Devereux, and Salvanes (2005), Oreopoulos, Page, and Stevens (2006) and Holmlund et al. (2011). Alternatively Carneiro, Meghir, and Parey (2013) use changes in school costs, and Maurin and McNally (2008) use modifications in examination modalities following the events of May 1968 in France. Different child's outcomes as various as years of schooling, grade repetition or post-compulsory school attendance have been studied. This last strategy using a natural experiment to instrument education is the one assessed in this thesis.

The third chapter of this thesis indeed analyses the possibility to exploit the events of May 1968 in France as a natural experiment to instrument education. The aim was initially to exploit these events to instrument parental education to be able to identify and estimate the causal link in the transmission of education from parents to children. Indeed during spring 1968, a student riot rapidly joined by a labor unions general strike paralyzed France and led to negotiations with the government. Regarding student demands, the evaluation conditions for high school and university examinations were a key issue as the school year had been disturbed for most students. In particular, the modalities of the French high school examination – the *baccalauréat* – were modified for the session 1968: they were simplified and led to a higher rate of success this year compared to adjacent years. Thus more high schoolers graduated this year and gained access to higher education. The strategy consisting in using the events of May 1968 in France to instrument education has been used by Maurin and McNally (2008) to assess both returns to education and intergenerational mobility. They find a high impact of the events on both earnings and children's school attendance. However such an important effect is surprising, as only a small proportion of the population was likely affected: only high schoolers taking the *baccalauréat* examination this year (only about a quarter of a birth cohort took it at this time), and who would not have succeeded without the modification of the examination modalities. Besides a new type of *baccalauréat* was created this year – the *baccalauréat technologique* – with a first session in 1969. Thus the events were not the only change in the educational structure at this time. Also Maurin and McNally (2008) use being born in 1949 as their instrument, since they do not dispose of the year students take their examinations, yet part of this birth cohort took the *baccalauréat* examination another year than 1968 and many students who took the examination in 1968 were born another year than 1949.

We implement a replication exercise and further investigate the validity of the instrument based on alternative data sets. It appears that being born in 1949 is not a convincing instrument for education. Using data from the Labor Force Survey (LFS) as the authors and the same empirical strategy, we find that the first-stage effect is very small and the instrument appears weak. We then reproduce the study on Census data to assess whether a larger data set would help produce better results, but the first-stage estimates are even smaller and the instrument fails at placebo tests. We suggest an alternative instrumental variable to verify whether the events of May 1968 qualify at all as a suitable natural experiment: having taken the *baccalauréat* examination in year 1968, information which is available in the Education-Training-Employment (FQP) survey and better represents the fact of having been affected by the events. The estimates are negligible and the instrument is weak again, and we conclude that the events of May 1968 actually did not increase the probability of obtaining the *baccalauréat* and pursue higher education studies. Even if the events of May 1968 increased the rate of success at the *baccalauréat* examination for year 1968, they had no substantial impact on the final level of education, as students had the opportunity to take the examination more than once. Thus a large part of students who obtained the *baccalauréat* in year 1968 thanks to the events would otherwise have repeated a grade and obtained it the next year. Finally, regardless of the choice of empirical procedure, we conclude that the events of May 1968 cannot be used to instrument education.

Chapter 1

Overview of intergenerational earnings mobility in Germany

1.1 Introduction

The extent of intergenerational transmission of socioeconomic status has interested economists for decades, as it reflects the impact of family background on inequality and thus the level of equality of opportunities of a society. The degree to which socioeconomic status is transmitted from one generation to the next indeed captures the impact family background can have on children's success in later life: the effect of factors independent from children's choices, talents and efforts on their future socioeconomic status. In other words, it represents to what extent childhood circumstances are reflected in adult life, or how children's success is determined by socioeconomic background.

The major theoretical model of income distribution and transmission across generations has been developed by Becker and Tomes (1979, 1986). The model of Solon (2004) is based on this work and formalizes the role of different mechanisms in the transmission: genetic heritability of income-generating characteristics and abilities, as well as human capital investment (possibly limited by credit constraints) and public policy (public provision of health care or education for instance). Empirical investigation of intergenerational economic mobility is developed in this framework and surveyed in particular in Solon (1999, 2002), Black and Devereux (2011) and Blanden (2013). As a measure of the degree of transmission, empirical studies have mostly analyzed the intergenerational elasticity (IGE), the regression coefficient relating child's log earnings to father's ones. A high IGE indicates a relatively rigid society, in which parental position in the income distribution and thus inequality is largely transmitted to the next generation. On the contrary, a low IGE reflects a more mobile society, in which children's economic success is largely independent of parental socioeconomic status.

Considering the existing literature, uncertainty about the extent of intergenerational economic mobility in Germany remains. The studies discussed here have exclusively employed data from the German Socio-Economic Panel (SOEP) – also used in this paper – and estimates range from 0.16 to 0.30 and from 0.18 to 0.36 in Couch and Dunn (1997, 1999) respectively, from 0.11 to 0.17 in Lillard (2001) and Couch and Lillard (2004), from 0.24 to 0.46 in Vogel (2006), from 0.28 to 0.37 in Eisenhauer and Pfeiffer (2008), around 0.26 and from 0.32 to 0.42 in Schnitzlein (2009, 2015) respectively. Thus results are highly sensitive to sampling and methodological procedures (Solon, 2002), specially concerning the method of treatment of biases. Methodological issues relative in particular to attenuation and life-cycle biases – investigated by Jenkins (1987), Solon (1989, 1992), Zimmerman (1992), Grawe (2006), Haider and Solon (2006) and Nybom and Stuhler (2011) – can indeed lead to severe misestimation of the persistence of inequalities, due to measurement errors of both generations' earnings. As lifetime measures of earnings are usually not available, current earnings have to be used instead. This errors-in-variables issue for parental earnings yields attenuation bias in the estimation of the IGE. Furthermore measurement error of both generations' earnings also leads to life-cycle bias since the association between current and permanent earnings evolves over the life-cycle.

The aim of this paper is to reassess intergenerational economic mobility in Germany. We first examine existing estimations of economic transmission in the recent literature about Germany. We also further investigate the impact of attenuation and life-cycle biases on the estimation of the IGE. In particular we estimate the evolution of the association between current and lifetime earnings over the life-cycle with data from the SOEP, the most frequently used data source in this literature. New estimates of the IGE in Germany are also presented, as well as the strong impact of uncorrected biases, from both sides measurement errors. The main analysis yields here an estimated IGE of 0.323. Not taking account of attenuation bias can lead to 30% lower estimates, whereas using averages of father's earnings over longer periods of time reduces this bias. Restricting son's age around 40 years old or adding interaction terms between son's age and father's earnings in the regression equation handles left-side life-cycle bias and thus leads to higher estimated IGE.

The remainder of this paper is organized as follows. Section 1.2 presents the estimation model and the issues faced when investigating IGE, as well as the estimation strategies implemented in the literature and here. Section 1.3 presents the methods and results found in the recent literature on intergenerational earnings mobility in Germany. Section 1.4 describes the data and sampling strategy. Section 1.5 presents the estimation results in terms of biases and IGE. Finally, Section 1.6 concludes.

1.2 Model and measurement

1.2.1 Estimation model

Among economists, intergenerational economic mobility is commonly measured by the link between the economic status of parents and children. The regression of son's earnings on father's one is estimated, using data expressed in logarithms for both generations, and the IGE is defined as the regression coefficient on father's income. A high elasticity suggests a strong impact of parental outcomes on children's ones, meaning little mobility inside the society. A low one on the contrary indicates a more mobile society and children's earnings less determined by parental ones.

The following regression model has typically been used to measure the IGE between fathers and sons, generally estimated by ordinary least squares (OLS):

$$lnY_{i,1} = \alpha + \beta lnY_{i,0} + \gamma Z_i + \epsilon_i. \tag{1.1}$$

Here $Y_{i,1}$ stands for the son's income[1] (generation 1) in family i, while $Y_{i,0}$ is the corresponding measure for the father (generation 0), and additional regressors Z_i can be included. The IGE β should not be given any causal interpretation: it includes all factors linked to income and transmitted across generations.[2]

1.2.2 Attenuation and life-cycle biases

One of the main difficulties in assessing the extent of economic persistence is that the regression equation should ideally be estimated using lifetime earnings measures for both sons and fathers. However in practice only short-run measures of earnings over a limited number of years are observed, since no data sets containing the required information are available. This exposes to two main types of biases. The first one is the attenuation bias (Altonji and Dunn, 1991; Solon, 1989, 1992; Zimmerman, 1992): using current earnings as proxies for permanent earnings for the explanatory variable leads to measurement error due to transitory fluctuations in measured earnings.

The second bias encountered in the estimation of the IGE is the life-cycle bias (Grawe, 2006; Haider and Solon, 2006). Indeed the previous attenuation bias is the only issue only if annual earnings as proxy for lifetime earnings follow the classical errors-in-variables model, for both generations. That is to say, there is no life-cycle

[1]This method can also apply to other measures of socioeconomic status.

[2]The intergenerational correlation (IGC) is an alternative measure of the degree of association between parents' and children's earnings, which takes into account potentially different distributions of earnings for each generation. In this paper the focus is mainly on IGE, this measure being widely used in the literature and in particular in most of articles discussed here.

bias only if the association between current and lifetime earnings does not evolve over the life-cycle and can be written:

$$lnY_{i,t} = lnY_i + \nu_{i,t}, \tag{1.2}$$

with $Y_{i,t}$ the current earnings observed for individual i in year t, Y_i the lifetime earnings and $\nu_{i,t}$ the measurement error. Otherwise, estimates of intergenerational earnings elasticity may be subject to inconsistency from both sides measurement errors.

Thus variation in the association between current and permanent earnings over the life-cycle also produces a bias in the estimation of the IGE, as reported by Jenkins (1987), and more recently by Haider and Solon (2006), Böhlmark and Lindquist (2006) and Nybom and Stuhler (2011). Haider and Solon (2006) provide evidence that indeed the slope coefficient λ_t in a regression of current on lifetime earnings systematically varies over the life-cycle and generally does not equal 1, generalizing the errors-in-variables model:

$$lnY_{i,t} = \lambda_t lnY_i + \nu_{i,t}. \tag{1.3}$$

For the estimation of the IGE, in the case of left-side measurement error, the probability limit of the estimated coefficient is then $\lambda_t\beta$ instead of β. In the case of right-side measurement error, this probability limit becomes $\theta_t\beta$ instead of β, with θ_t also called "reliability ratio", the slope coefficient in the "reverse regression" of lifetime earnings on current earnings. According to the estimations of Haider and Solon (2006), Böhlmark and Lindquist (2006) and Brenner (2010) for the US, Sweden and Germany respectively, λ_t ranges from around 0 to 1.3 depending on the age at which son's earnings are observed and θ_t ranges from around 0 to 0.8 depending on father's age.

Intuitively, as explained in Haider and Solon (2006), workers with high lifetime earnings tend to also have high earnings growth rates. Thus if sons are observed shortly after entering the labor market and fathers shortly before leaving it – which is usually the case – the difference in lifetime earnings of skilled and unskilled sons (fathers) is under- (over)estimated (Reville, 1995). Therefore the age at which earnings are observed for both generations is an important issue and can lead to severe under- or overestimation of the IGE.

1.2.3 Estimation strategies

In order to decrease the magnitude of the attenuation bias consecutive to right-side measurement error, parental earnings can be averaged over several years to reduce the variance of the noise relative to the signal, as explained in Solon (1989, 1992) and Zimmerman (1992) and implemented in most of the recent literature. However according to Mazumder (2005), IGE estimated with father's earnings averaged over only five years could still be biased from around 30% in the US. The estimate $\hat{\beta}$ of β would then be biased toward 0 by an attenuation factor θ_t of 0.69. If the average covers 10 years, this ratio raises to 0.79. Earnings observations for 25 years would be necessary to reach a reliability ratio of 0.90.

Parental earnings can alternatively be instrumented by parents' education level and/or occupational status. Then the estimated IGE is unbiased only if the instrument is uncorrelated with son's earnings or perfectly correlated with father's ones. Because father's education and occupation have a positive impact on son's earnings, the IGE is presumably overestimated with instrumental variables (IV) estimation (Solon, 1992; Nicoletti and Ermisch, 2008). However the IGE being on the contrary underestimated with OLS estimation if attenuation bias is only reduced and not completely removed, the real IGE should lie in between. Björklund and Jäntti (1997) for instance find empirical results estimated at 0.39 and 0.52 using OLS and IV estimation respectively, for the case of the US. Nonetheless Nybom and Stuhler (2011) argue that IV estimates are not necessarily upper bounds for the IGE and that IV estimation is a less satisfactory way to handle attenuation bias in the estimation of the IGE than averaging father's earnings over several years (even if this method only reduces and do not completely correct the bias).

To handle left-side life-cycle bias, sons can be observed at ages for which the coefficient λ_t is around 1. Haider and Solon (2006) precisely estimate the association between annual and (the present value of) lifetime earnings over the life-cycle and find current earnings to be acceptable proxies for lifetime earnings "between the early thirties and the mid-forties" for the US, using nearly career-long Social Security earnings histories.[3] Böhlmark and Lindquist (2006) and Brenner (2010) make similar recommendations for Sweden and Germany respectively.

Based on Grawe (2006) and Haider and Solon (2006) explaining that the classical errors-in-variables model does not apply for the estimation of the IGE, Lee and Solon (2009) implement an alternative strategy to deal with life-cycle bias: they control not only for child's age but also for the interaction between child's age and parental income in the regression of child's log income on parental ones, to account for the systematic heterogeneity across individuals in income growth's rates over the life-cycle.[4]

Nybom and Stuhler (2011) highlight the limits of current empirical strategies used to correct life-cycle bias and based on the generalized errors-in-variables model. In particular they object that the coefficient λ_t "is merely a population parameter that reflects how differences in annual income and differences in lifetime income relate *on average in the population.* Individuals will nevertheless deviate from this average relationship, so that their annual income still over- or understates their life-

[3] Haider and Solon (2006) have access to labor earnings data for the period 1951-1991 for people born between 1931 and 1933, i.e. aged 18-20 at the beginning of the period and 58-60 at the end. They restrict their sample to workers with earnings available in at least 10 years and obtain a sample of 821 individuals.

[4] Lee and Solon (2009) use PSID family income data for sons and daughters born between 1952 and 1975 and observed from age 25 to 48, i.e. for the period 1977-2000. Parental family income is calculated as the average over the three years when the child was 15-17 years old. This strategy yields samples of 1228 sons and 1308 daughters, with almost 10 observations per individual on average.

time income." They point out that for the estimation of the IGE, such deviations should not depend on family background, whereas it is likely the case. They evaluate at around 20% the remaining bias after correcting the estimates for left-side measurement error, using nearly complete Swedish income series,[5] and provide recommendations to reduce it. In particular they advise to average not only father's, but also son's earnings information in order to partially correct the left-side measurement problem.

In this paper, we first estimate λ_t and θ_t in order to investigate the extent of the biases encountered in the estimation of the IGE in Germany, and the ages of sons and fathers at which earnings should be observed to reduce them.

For the investigation of the IGE, our main estimation strategy consists in the estimation of equation (1.1) with age and age squared for both generations as additional regressors Z_i, in order to control for the effects of changes in earnings during a career. We average both father's and son's earnings over time, as recommended in Nybom and Stuhler (2011).

To assess the effect of attenuation bias due to right-side measurement error, we average father's earnings over different periods of time, following Mazumder (2005). Concerning the investigation of the bias due to left-side measurement error, we implement two strategies. First we restrict the age range at which we observe son's earnings to values for which λ_t is supposed to be around 1. Alternatively we implement the method of Lee and Solon (2009), consisting in the addition of an interaction term between son's age and father's earnings in the estimation of the IGE.

[5] Nybom and Stuhler (2011) use Swedish tax registry data for the years 1960-2007 for sons born 1955-1957 and their biological fathers. They restrict their sample to fathers and sons reporting total income in at least 10 years, which leads to a sample of 3,504 father-son pairs with son's income observed between age 22 and 50 and father's income observed between age 33 and 65.

1.3 Recent literature on intergenerational earnings mobility in Germany

For the case of Germany, and in particular in all studies discussed here, the question of intergenerational economic mobility has been assessed relying on data from the German Socio-Economic Panel (SOEP) and mainly estimating intergenerational elasticities or correlations. The summary Table 1.1 reports recent results obtained for Germany in the literature and the way authors addressed the issues of attenuation and life-cycle biases.

Table 1.1: Recent results and treatment of biases in the German literature

Authors	**Year**	**Est.**	**Sampling strategy**	**Treatment of attenuation bias**	**Treatment of life-cycle bias**
Couch	1997	0.16	s. from age 18	up to 6 year average (1)	
& Dunn		0.30	s. from age 25	up to 6 year average (1)	s.' age restriction
	1999	0.18	s. from age 22	min 3 years average	
		0.24	s. from age 22	min 4 years average	
		0.19	s. from age 22	min 5 years average	
		0.28	s. from age 25	min 3 years average	s.' age restriction
		0.36	s. from age 25	min 4 years average	s.' age restriction
		0.36	s. from age 25	min 5 years average	s.' age restriction
Lillard	2001	0.11	18-60	average	
Couch	2004	0.13	s. from age 18	average	
& Lillard		0.17	s. from age 25	average	s.' age restriction
Vogel	2006	0.24	25-60	min 5 years	no groups
		0.31	25-60	min 5 years	skill groups (2)
		0.41	25-60	min 10 years	no groups
		0.45	25-60	min 10 years	skill groups (2)
Eisenhauer	2008	0.20	30-50	single-year	(3)
& Pfeiffer		0.28	30-50	5 year averages	(3)
		0.36	f. 30-65	5 year averages	(3), f. up to 65 yo
		0.24	s. 20-50	5 year averages	(3), s. from 20 yo
		0.37	30-50	IV estimation	(3)
Schnitzlein	2009	0.26	f. 30-55, s. 30-40	min 5 year average	
	2015	0.33	f. 35-55, s. 35-42	min 5 year average	s.' age restriction

Notes: Abbreviations: s. for sons, f. for fathers
(1) mean of all estimations from single-year and up to six-year averages of parental observations
(2) four skill groups with different wage growth ; (3) a unique pair of father/son observations is kept: with the smallest father/son age difference to observe them at the most similar stage of life-cycle possible

1.3.1 Couch and Dunn (1997, 1999)

Little attention had been paid to intergenerational economic transmission in Germany before the article of Couch and Dunn (1997) and their comparison of intergenerational mobility in Germany and in the US in terms of earnings, work hours and education. For the case of Germany, they use data from the 1984 to 1989 surveys of the SOEP to observe annual labor earnings for sons and their fathers, and average these earnings over the six survey years for sons. For fathers, the economic outcome is alternatively defined as the average over one year, two years, and so on up to six years. Excluding observations for years during which sons were in school or fathers in school or retired, they obtain a sample with on average 22.8 years old sons and 51.0 years old fathers.

They find extremely low estimates of intergenerational earnings correlation in their main analysis: around 0.16 for both Germany and the US, when they take the average of all estimates obtained with parental outcome defined as single-year observation and two to six-year averages of observations.[6] However the large difference between these results and those of Solon (1992) for instance can be explained by different sampling procedures. In particular Couch and Dunn (1997) include sons from age 18 as long as they are out of school – versus 25 in Solon (1992) – and thus face a severe underestimation bias due to life-cycle effects. When Couch and Dunn (1997) observe sons only from age 25, the correlation rises to around 0.30 for Germany (and 0.26 for the US) since it yields partial correction of the bias.

Couch and Dunn (1999) add the United Kingdom to their comparison of intergenerational mobility in a subsequent paper and reassess their previous results with more recent data. For Germany, they use the waves 1991-1995 of the SOEP and average both son's and father's available earnings observations over this period. Restricting the analysis to sons from age 22 and fathers until age 65 (again out of school

[6] Couch and Dunn (1997) find an estimated IGC of 0.121 – and an estimated IGE of 0.112 – when they use an average of all available observations over the six survey years as the outcome for both generations, with a sample of 272 father-son pairs.

and retirement), they obtain a sample of 388 father-son pairs with sons aged 27 and fathers 54 on average, reporting respectively 3.5 and 4.2 available earnings observation on average. They find estimated elasticities (resp. correlations) ranging from 0.18 to 0.24 (resp. 0.21 to 0.28) when imposing minimum 3 to 5 available earnings observations for both sons and fathers. The estimated elasticities (resp. correlations) range from 0.28 to 0.36 (resp. 0.33 to 0.40) when the sample is restricted to sons from age 25.[7]

1.3.2 Lillard (2001), Couch and Lillard (2004)

Lillard (2001) and Couch and Lillard (2004) also compare the extent of intergenerational earnings persistence in Germany and in the US. For Germany, they use data from the waves 1985-1998 of the SOEP. They select men reporting earnings between the ages of 18 and 60, and as in Couch and Dunn (1997) they also exclude earnings observations for men who were in school, retired or not in the labor force. Couch and Lillard (2004) then raise the lower age restriction to 25, as in most of the literature. Earnings are calculated as the average of all available annual observations.

In Lillard (2001), this sample strategy yields a sample of 1,061 father-son pairs, with sons aged 26 and reporting 6 earnings observations on average, and fathers aged 52 and reporting 8 earnings observations on average. In Couch and Lillard (2004), the sample including sons aged 18 and older consists of 657 father-son pairs[8], with fathers aged 50 and sons 26 on average, and earnings computed from 11 years for fathers and 8 years for sons on average. In the sample of 549 father-sons pairs including sons aged 25 and older, fathers are 51 years old and sons 29 on average, and earnings are computed from 11 years and 6 years on average, for fathers and sons respectively.

[7] The sample size is reduced to 218, 144 and 102 father-son pairs when 3, 4 and 5 available observations are required for both sons and fathers, for the sample with sons from age 22. For the sample with sons from age 25, the corresponding sample sizes are 150, 104 and 64.

[8] Variation in the sampling strategy can yield the sample size difference. Couch and Lillard (2004) only include men who worked at least 850 hours in one of the years with reported earnings, which can lead to a smaller sample as in Lillard (2001), where such restriction is not imposed.

Elasticities are estimated at 0.11 in Lillard (2001), and in Couch and Lillard (2004) at 0.13 when sons from age 18 are selected, 0.17 when the sample is restricted to sons from age 25. As in Couch and Dunn (1997), the estimates are extremely low, in particular when sons from age 18 are included in the sample. The estimate is again higher when the sample is restricted to sons at older ages, however the increase is much smaller than in Couch and Dunn (1997, 1999) and the estimate remains lower than expected.

1.3.3 Vogel (2006)

Vogel (2006) reassesses the comparability of intergenerational economic mobility in Germany and in the US, further investigating life-cycle issues discussed by Jenkins (1987) and more recently Grawe (2006) and Haider and Solon (2006), and presenting an estimation strategy to correct it. For Germany, he uses data from the waves 1984-2005 of the SOEP to observe annual labor earnings of fathers and sons between 25 and 60 years old, when they are available for minimum five years. This yields a sample of 525 sons from 421 fathers, respectively 30.4 and 50.4 years old on average.

In order to estimate the IGE, Vogel (2006) implements a two-step estimation strategy. First, he estimates life-cycle earnings profiles, based on an income-generating function linearly increasing in time and with a quadratic age effect. Since earnings of sons (resp. fathers) are mostly observed in the early (resp. late) stages of their careers but not at he end (resp. beginning), the assumption is made that earnings profiles of sons and fathers are identical. This allows to use observations from all men aged 25 to 60 and with at least five available earnings observations, yields a large data set of 5,089 individuals and limits measurement error and thus attenuation bias (see Section 1.2.2). Then lifetime earnings of sons and fathers are computed based on the previous estimation and used to obtain an estimate of the IGE.

A seen in Section 1.2.2, a comparison of current earnings of workers with high and low lifetime earnings tends to under- (resp over-)estimate their gap in lifetime earnings at early (resp. late) stages of the career. Therefore to correct life-cycle bias, Vogel (2006) considers four types of workers, allowing different wage growth profiles. Estimating income-generating functions separately for these skill groups indeed reveals very different earnings growth rates, greater for higher educated individuals. The benchmark estimation leads to an estimated IGE of 0.24 in Germany, much lower than the estimate of 0.31, when differences in wage growth rates are taken into account.[9]

Vogel (2006) presents an alternative method to handle life-cycle bias to the generalized errors-in-variables model and finds a strong impact of this bias on German estimates of the IGE. However Nybom and Stuhler (2011) explain that even within educational groups, other determinants of income linked to family background can lead to deviation from the mean income growth rate (see Section 1.2.3), and thus that this strategy does not eradicate life-cycle bias as Vogel (2006) argues, even if it improves the estimation.

1.3.4 Eisenhauer and Pfeiffer (2008)

Eisenhauer and Pfeiffer (2008) estimate the IGE in Germany using the waves 1984 -2006 of the SOEP. Their measure of economic status is real monthly earnings of full-time employed workers between 30 and 50 years old, with only the oldest sons included into the sample. Moving averages of earnings observations over five years are implemented for fathers, to reduce the attenuation bias. IV estimation is also implemented with parental years of education as instrument for long-run parental status, as in Solon (1992) and Dearden et al. (1997).

To reduce life-cycle bias, Eisenhauer and Pfeiffer (2008) compute a sampling procedure leading to observe father-son pairs at the most similar stage in their lives as possible. Therefore father's and son's observations are matched in all possible combinations and a unique pair is identified: the one with the smallest absolute age

[9]The same estimations when earnings observations are required for at least ten years are 0.41 and 0.45 respectively, attenuation bias being further reduced.

difference between father and son (and then associated with the lowest father age, if needed). This yields a sample of 180 father-son pairs, with 35.7 years old sons and 44.4 years old fathers on average.

The main analysis leads to an estimated elasticity of 0.28. Investigating biases, the estimated IGE increases with the number of years averaged and when IV estimation is implemented (IGE estimated at 0.37), as attenuation bias declines. It also increases when the upper age limit for fathers is raised, presumably due to a reduction of transitory fluctuations, and decreases when the age requirement for sons is lowered, life-cycle bias rising. This leads to estimated IGE of 0.36 when the upper age-limit for fathers is 65 and 0.24 when the downer age-limit for sons is 20 years old.

Eisenhauer and Pfeiffer (2008) suggest a point estimate of the IGE in Germany at one third. However their strategy is to observe fathers and sons at the closest stage in life as possible, which contradicts the results of Brenner (2010) who recommends to observe older fathers than sons. In pratice their sample follow the suggested age ranges. Then they only investigate the extent of life-cycle bias by changing age restrictions for sons and fathers. One could argue that part of the effect can be driven by an evolving IGE across cohorts, and not only by a reduction of the bias.[10]

1.3.5 Schnitzlein (2009, 2015)

Schnitzlein (2009) uses the classical OLS estimation of IGE to measure the extent of intergenerational mobility in Germany, with the waves 1984-2005 of the SOEP. He handles the issue of attenuation bias by averaging labor earnings observations for fathers over minimum five years. Only one available observation is required for children (sons and daughters), but all earnings observations are also averaged. Fathers are observed between age 30 and 55, and children between age 30 and 40.

[10] See among others for trends in the IGE Hertz (2007), Lee and Solon (2009) and Aaronson and Mazumder (2008) for the US, Nicoletti and Ermisch (2008) and Erikson and Goldthorpe (2010) for Britain, Lefranc (2011) for France and Björklund et al. (2009) for Sweden.

Additionally, a lower income limit of 1,200 euros per year is implemented for both generations. For sons, this leads to a sample of 439 father-son pairs, with 34.3 years old sons and 47.1 years old fathers on average. The IGE is estimated at 0.26.

In a subsequent paper, Schnitzlein (2015) conducts a cross-country comparision of levels of intergenerational earnings mobility in Germany and in the US. For the German part, he observes annual labor earnings from the waves 1984-1993 of the SOEP for fathers and 1997-2011 for sons. Fathers are again observed at age 30 to 55 but sons when they are between 35 and 42 years old, based on the findings of Haider and Solon (2006), to limit life-cycle bias. Here earnings observations have to be available for more than five years for fathers and only one for sons, and are averaged. Different lower annual earnings limits are alternatively computed: 1,200 euros in the main estimation, then 4,800 and 9,600 euros.

This yields a sample of 408 father-son pairs, with 37.4 years old sons and 47.3 years old fathers on average. IGE is estimated at 0.32 for Germany (with the lower annual income limit of 1,200 euros, results varying substantially with income restrictions, up to 0.42). In line with the findings of Couch and Dunn (1997) but contrary to those of Vogel (2006), Schnitzlein (2015) finds similar estimates of the IGE in Germany and in the US.

It seems that estimates of intergenerational mobility are not very robust against differences in sampling procedure and are especially highly sensitive to the treatment of attenuation and life-cycle biases. In the papers presented here the methods to handle these biases are various, particularly for life-cycle issues. However the IGE in Germany seems to be consistently estimated around 0.3 when attenuation and life-cycle biases are taken into account, even if these biases may remain partly uncorrected.

1.4 Data

1.4.1 SOEP data and main variables

This paper uses data from the German Socio-Economic Panel (SOEP) (Wagner, Frick, and Schupp, 2007), a nationally representative household survey started in 1984 and conducted annually. All adult members of each household are part of the survey and followed as long as possible and in other locations, in particular when children leave parental home and form their own households. It is thus possible to relate children's economic status as adults to their parents' status. However as the survey only started in 1984 the panel is still relatively short, and as children have to still live in the household when their parents are interviewed to become a member of the survey, sons who left late the parental home are potentially overrepresented. This would lead to keep more highly educated sons who therefore achieve better economic success and thus to underestimate the IGE.

In this study, the SOEP panel is separated in two equal parts of 15 waves each: waves 1984 to 1998 are used to observe fathers and waves 1999 to 2013 to observe sons. We choose to use as variable of interest the individual annual labor earnings[11] from the SOEP and included in the Cross-National Equivalent File (CNEF) (Frick, Jenkins, Lillard, Lipps, and Wooden, 2007). All earnings information are deflated by the Consumer Price Index (CPI), base year being 2005.

1.4.2 Sample selection and descriptive statistics

In a first attempt to reproduce the results of recent literature, we implement the same sampling procedure as in Schnitzlein (2015). A very similar sample is obtained, reported in Table 1.2[12].

[11] The variable includes wages and salary from all employment including training, primary and secondary jobs, and self-employment, plus income from bonuses, over-time, and profit-sharing.

[12] See Table A.1 (Full SOEP Sample) in the Additional Supporting Information which can be found in the on line version of the article at the publisher's website for the corresponding information obtained in Schnitzlein (2015).

Table 1.2: Descriptive statistics with Schnitzlein's empirical strategy

	Mean/Median	**Std Dev.**	**Min**	**Max**
son's earnings	37,911	22,563	2,370	227,625
father's earnings	32,063	17,516	9,129	145,762
son's age	37.41	1.30	35	41
father's age	47.33	4.19	33.5	52.5
son's number of observations	5.40	2.41	1	8
father's number of observations	9.16	1.27	6	10

Notes: 408 observations; Earnings in 2005 euros; Median of earnings, mean for all other variables

Then for the main analysis, we observe sampled fathers between 30 and 55 years old, as in Schnitzlein (2009, 2015), and whose earnings observations have to be available for at least five years, following the recommendations of Solon (1989, 1992). We compute the average of these earnings observations to reduce attenuation bias, as seen in Section 1.2.2. We observe sons alternatively when they are aged 30 to 55 or 35 to 42, to investigate the extent of potential life-cycle bias, since restrincting sons' age range should reduce it, as explained in Haider and Solon (2006) and confirmed in Böhlmark and Lindquist (2006) and Brenner (2010). We also average their earnings observations overtime to reduce potential measurement error, following Nybom and Stuhler (2011) as seen in Section 1.2.2, without any restriction on the minimum number of available annual observations. If more than one son are matched to a father, we use all father-son pairs (with standard errors corrected for family clustering). In order to investigate possible sample homogeneity arising from the potential overrepresentation of sons who left late the parental home, we also estimate IGE computed with only sons still young when fathers are interviewed: with an alternative sample restricted to sons who were younger than 18 years old in year 1984.

The main sample procedure yields a sample of 652 father-son pairs with sons observed between 30 and 55 years old, 448 pairs when sons are only observed between 35 and 42 years old. When restricting the analysis to sons younger than 18 years old in 1984, the sample is reduced to 202 father-son pairs. The main descriptive statistics are reported in Table 1.3 and show sons are observed at age either 35 or 37.5 on average, depending on the sampling strategy. Fathers are observed around 47-48 years old on average. Averaged earnings are computed using more than five annual earnings observations for sons and more than ten for fathers, with a required minimum of five available observations.

Table 1.3: Descriptive statistics for the main analysis

	Mean	**Std. Dev.**	**Min.**	**Max.**
Sons aged 30 to 55, father's earnings averaged over at least 5 years - 652 pairs				
sons' earnings	37,492	22,488	524	248,740
fathers' earnings	37,165	19,504	6,359	230,479
sons' age	35.20	3.85	30	48.5
fathers' age	47.02	4.72	32.5	53
sons' number of observations	7.79	4.83	1	15
fathers' number of observations	11.21	3.26	5	15
Sons aged 35 to 42, father's earnings averaged over at least 5 years - 448 pairs				
sons' earnings	42,136	23,579	418	227,625
fathers' earnings	37,045	18,217	9,129	155,788
sons' age	37.64	1.40	35	42
fathers' age	48.14	3.80	34.56	53
sons' number of observations	5.45	2.39	1	8
fathers' number of observations	10.88	3.23	5	15

Note: All earnings in 2005 euros

1.5 Results

As a first observation, we are able to accurately reproduce the results found in Schnitzlein (2015), when implementing the same empirical strategy. Indeed the IGE is estimated at 0.318 (standard error: 0.072) on a sample of 408 father-son pairs in Schnitzlein (2015) and our estimation is 0.325 (standard error: 0.071) also on a sample of 408 father-son pairs.

1.5.1 Estimation of biases

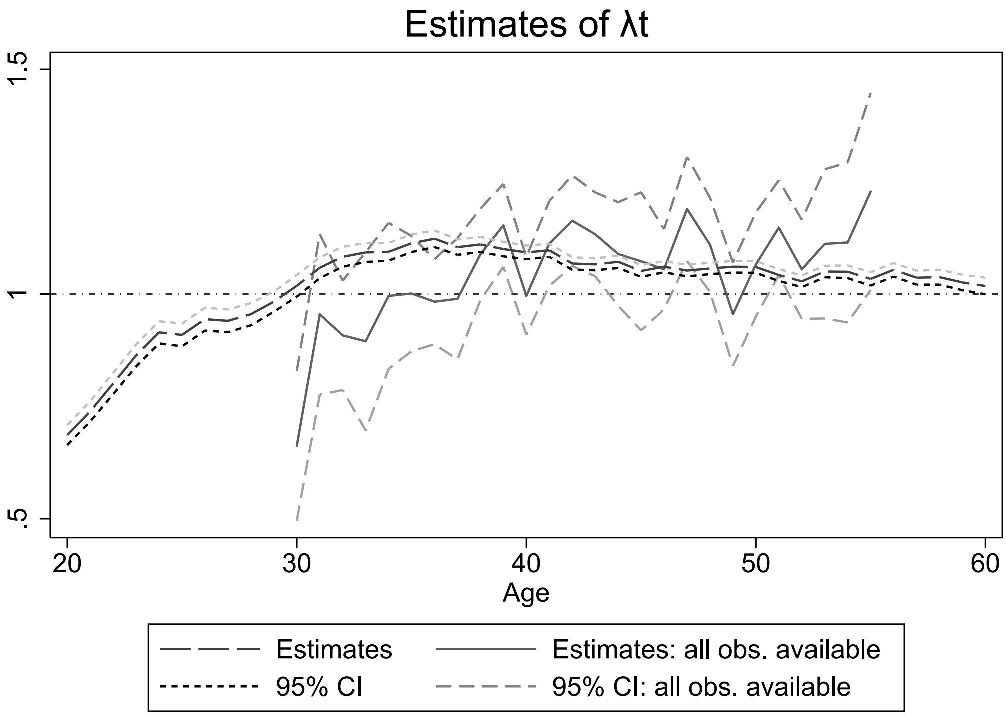
To investigate the bias induced by both sides measurement errors in terms of lifetime earnings, and to assess the age at which sons' and fathers' earnings should be observed to reduce it, we estimate the coefficients λ_t and θ_t defined in Section 1.2.2. Using 30 waves of the SOEP (1984 to 2013), we regress log annual earnings for each age from 20 to 60 on the log of the present discounted value of lifetime earnings to estimate λ_t, following the method of Haider and Solon (2006).[13] Using the same strategy, we also compute the "reverse regressions" of the log of the present discounted value of lifetime earnings on log annual earnings to estimate θ_t.

We compute here lifetime earnings with all available earnings information, but with no restriction on the minimum number of available observations. Ideally we would need career-long earnings history as in Nybom and Stuhler (2011), which is nearly the case in Haider and Solon (2006), Böhlmark and Lindquist (2006) and Brenner (2010), but not possible with our data. So additionally, we estimate the coefficients for the restricted group of individuals with all 26 observations available from age 30 to 55, assuming this should yield a good indicator of lifetime earnings. These coefficients λ_t and θ_t, for the whole and the restricted groups, are represented in Figure 1.1 and 1.2 respectively.

Brenner (2010) estimates λ_t and θ_t also for Germany, but uses data from the *Vollendete Versichertenleben* of the Research Data Centre of the German Statutory Pension Insurance, and not from the SOEP which is more widely used in the intergenerational mobility literature. He selects individuals born between 1939 and 1944, observes their earnings between age 19 and 59, and restricts the sample to individuals with at least 10 available income observations. Thus we observe individuals essentially for the same age range, but use different restrictions for the minimum number of available information. Furthermore his economic outcome is gross annual income subject to social insurance contribution, which comes with two drawbacks we do not have: the information is censored as it is only reported up to a contribution ceiling and neither civil servants nor most of self-employed are represented.

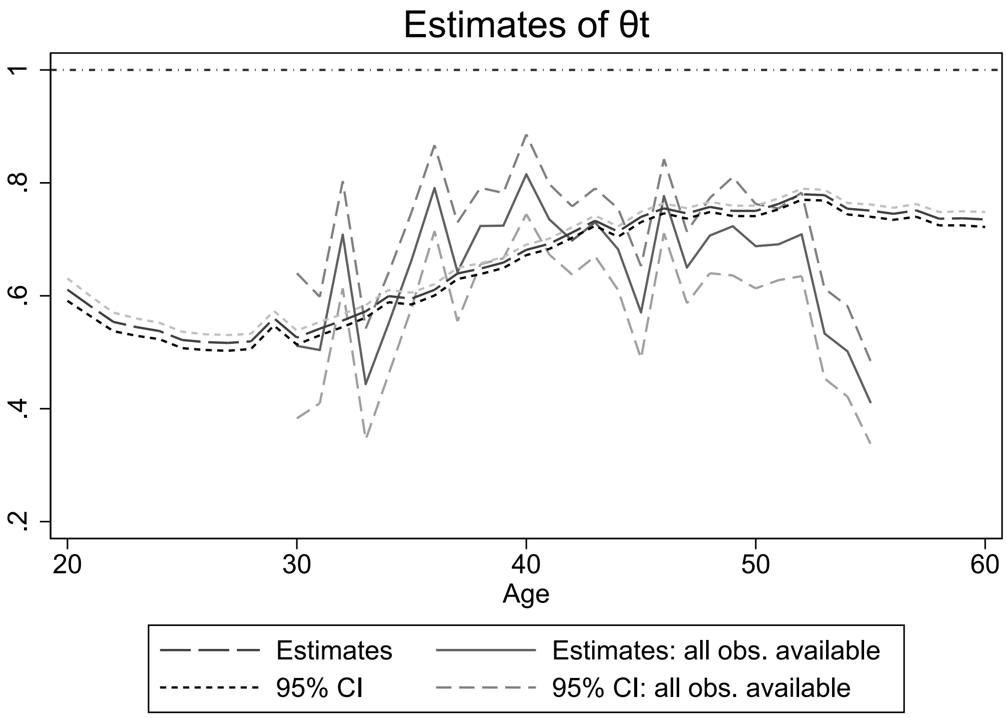
[13]The present discounted value of lifetime earnings is calculated using the CPI to convert each year's nominal to real earnings, and assuming a constant real interest rate of 2 percent to maintain comparability to the literature.

Figure 1.1: Estimates of λ_t for all individuals aged 20 to 60 and restricted to individuals aged 30 to 55 with all 26 earnings observations available



The graphic representation of λ_t in Figure 1.1, for all individuals aged 20 to 60 and without restriction on the number of available earnings observations, shows a substantial bias if earnings are observed in the early stage of the career. This bias then decreases until around age 30, when λ_t reaches 1. After this, the bias slightly increases during the early thirties with λ_t exceeding 1, but remains low and then slowly decreases without falling again below 1, contrary to the findings of Haider and Solon (2006) for the US. When we restrict the sample to individuals observed between age 30 and 55 and with all earnings observations available for this period, we estimate λ_t below 1 until a few years later in the mid-thirties, near 1 until the forties and then above 1, without the previous decreasing trend. These results on the restricted sample are very similar to those of Brenner (2010), which are also close to the results of Böhlmark and Lindquist (2006) for Sweden. This depiction of λ_t confirms observing son's earnings at young ages leads to a strong underestimation of the IGE due to left-side life-cycle bias, and it seems the IGE can be satisfactorily estimated if sons are observed around 35 or 37.5 years old on average, as it is the case in this study (depending on whether sons are observed from 30 to 55 years old or from 35 to 42 years old).

Figure 1.2: Estimates of θ_t for all individuals aged 20 to 60 and restricted to individuals aged 30 to 55 with all 26 earnings observations available



Concerning right-size measurement error, as depicted in Figure 1.2, we always estimate θ_t below 1, for both samples, as in Brenner (2010), Haider and Solon (2006) and Böhlmark and Lindquist (2006). For all individuals aged 20 to 60, θ_t slightly decreases until the late twenties, increases until the mid-forties then remains stable until the early fifties, and finally shows a slow decreasing trend. When we restrict the sample to individuals with career-long earnings observations, θ_t also seems to increase until the forties, then remains almost stable before falling in the early fifties. Again our estimations of θ_t for the restricted sample are relatively close to the one found in the literature. They confirm a severe bias if fathers are observed early in their career or at its very end, and show fathers should be observed later than sons for the estimation of the IGE. Observing them around 47-48 years old on average, as in this paper, appears adequate. Some bias however still remains, as we estimate θ_t around 0.75 at these ages.

1.5.2 Estimation of the IGE

In the first part of the estimation of the IGE, based on the empirical strategy presented in Sections 1.2 and 1.4, we observe earnings for fathers and sons between 30 and 55 years old. We investigate the magnitude of the attenuation bias by averaging father's earnings over shorter or longer periods of time, since the potential bias might be high, as highlighted in Mazumder (2005). We present in Table 1.4 the estimated IGE without any restriction on the number of available earnings observations and then averaging over minimum 5 and minimum 10 years. Mazumder (2005) suggested to average parental earnings over even longer periods, but the SOEP is not a long enough survey for us to do so.

Table 1.4: Estimated IGE with average of father's earnings over different periods of time - unbalanced panel

	no min.	min. 5 years	min. 10 years
β	0.194 ***	0.286 ***	0.337 ***
std err.	(0.049)	(0.061)	(0.075)
fathers' age	47.87	47.02	45.79
fathers' nb of obs.	9.43	11.21	13.17
obs.	818	652	434

Significance levels : $*$: 10% $**$: 5% $***$: 1%

With no minimum on the number of earnings observations, the IGE is estimated at only 0.194, with a sample of 818 father-son pairs.[14] This estimate is much lower than the one obtained when we average father's earnings over minimum five years: 0.286, obtained from 652 father-son pairs. If father's earnings observations are

[14] The corresponding IGC amounts 0.324, with standard deviations of log earnings of 0.677 and 0.406 for sons and fathers respectively. Further only estimated IGE are reported, IGC being always higher due to a higher dispersion of sons' earnings distribution than fathers' one, but yielding similar conclusions.

averaged over minimum ten years, the estimated IGE amounts to 0.337, from 434 father-son pairs.[15]

These results underline the importance of removing or at least reducing attenuation bias due to right-side measurement error, to avoid serious underestimation of the IGE. Using more or less father's earnings observations yields estimates depicting completely different images of intergenerational mobility, even if more than 9 years on average are already used without any required minimum.

However the large differences between these estimates can come from the different samples used and are probably also driven by the differences in father's age, not only in the number of earnings observations averaged. Therefore we keep the sample of 434 individuals with father's earnings known for at least 10 years and reestimate IGE with only one father's earnings observation and with father's earnings averaged over 5 and 10 years. We choose the observations used so that father's age remains as stable as possible, that is close to 45.79 years old. The results reported in Table 1.5 still show the importance of handling attenuation bias, the IGE being estimated at 0.224, 0.292 and 0.332 with the three different specifications. In the rest of the paper, we average father's earnings observations over minimum five years to reduce the magnitude of attenuation bias, following most of the literature on this subject.

Table 1.5: Estimated IGE with average of father's earnings over different periods of time - balanced panel

	1 year	5 years	10 years
β	0.224 ***	0.292 ***	0.332 ***
std err.	(0.060)	(0.074)	(0.073)
fathers' age	45.54	45.55	45.71
obs.	434	434	434

Significance levels : $*$: 10% $**$: 5% $***$: 1%

As seen previously, notably through our estimation of λ_t in Section 1.5.1, observing sons' earnings when they are not in a certain range of ages in midlife can lead

[15] When an average over all fifteen years is used, the estimated IGE is no longer significant and falls to 0.209 with a sample reduced to only 179 father-son pairs. However fathers are less than 44 years old on average in this sample, and around 46, 47 and 48 years old with a minimum of 10 years, 5 years, and no minimum, respectively.

to life-cycle bias. In order to investigate the magnitude of this bias, we compare the IGE resulting from the previous analysis to an IGE estimated with sons observed from age 35 to 42. This age restriction, also applied in Schnitzlein (2015), is chosen according to the results of Haider and Solon (2006) and in line with ours. With this sampling strategy we estimate the IGE at 0.323, as reported in Table 1.6, which is in range with the recent literature on intergenerational economic transmission in Germany, and very close to the one found using the empirical strategy of Schnitzlein (2015), estimated at 0.325.

Table 1.6: Estimated IGE with sons observed at different ages and the addition of an interaction term between son's age and father's earnings (coefficient δ)

	sons at age 30 to 55	sons at age 35 to 42	with interaction term
β	0.286 ***	0.323 ***	0.356 ***
std err.	(0.061)	(0.069)	(0.089)
δ			0.013
std err.			(0.016)
obs.	652	448	652

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Lee and Solon (2009) suggest another way to treat life-cycle bias. They add interaction terms of polynomials of son's age and father's earnings in the regression equation. Here we include as regressor an interaction of son's age with father's earnings. We again observe sons between 30 and 55 years old. Because of likely multicollinearity between ages and the interaction term, we remove son's and father's age squared from the equation. This alternative procedure yields an estimated IGE of 0.356, as reported in Table 1.6, thus even higher (even if not significantly higher) than the one found with the age restriction. However the interaction term coefficient is not significant.

As a last concern about the sampling procedure, the SOEP started in 1984 and is a rather short data set, with only 30 waves currently available. Moreover children had to live in the parental household when their parents were interviewed to be part of the survey as they formed their own household. Thus some of the sampled sons probably stayed in the parental home until an advanced age.

To investigate this possible sample homogeneity, we estimate the IGE with a sample excluding sons older than 18 years old in 1984 (using sons aged 30 to 55 years old, to avoid a too small sample and a not significant estimate). This procedure yields a smaller sample of only 202 father-son pairs and the results are reported in Table 1.7. Excluding sons too old in 1984 increases the value of the estimated IGE from 0.286 to 0.339. This supports the idea that the homogeneity bias arising from the overrepresentation of high achieving sons would lead to an underestimation of the IGE. However our estimates are not significantly different and it is complicated to draw more than caution conclusions due to the small size of the sample.

Table 1.7: Estimated IGE without sons who stayed late in the parental household

	all sons	younger than 18 in 1984
β	0.286 ***	0.339 ***
std err.	(0.061)	(0.115)
fathers' age	47.02	45.46
fathers' nb of obs.	11.21	11.81
obs.	652	202

Significance levels : $*$: 10% $**$: 5% $***$: 1%

1.6 Conclusion

This paper investigates intergenerational earnings transmission in Germany, using data from the SOEP and estimating the IGE. Uncertainty indeed remains about the extent of mobility in this country, as well as concerning a satisfactory way to choose sampling strategy and handle biases. In the main analysis, earnings observations are collected for sons between 35 and 42 years old to reduce left-side life-cycle bias, whereas father's earnings information, observed at ages 30 to 55, are averaged over minimum five years in order to reduce attenuation bias due to right-side measurement error. This leads to an estimated IGE of 0.323 for Germany. This estimate is in range with the recent literature, and in particular very close to the estimated IGE obtained by Schnitzlein (2015) with a similar estimation strategy.

The magnitude of attenuation and life-cycle biases is also examined. First graphically, a depiction of the coefficients λ_t and θ_t representing left-side and right-side life-cycle biases respectively, support the observation of sons aged around 38 and fathers around 48. Then, focusing on attenuation bias due to right-side measurement error, father's earnings are averaged over different periods of time, and serious potential attenuation bias is revealed. Using a single father's earnings observation or an average over too few years yields severe underestimation of the IGE. Concerning left-side life-cycle bias, reducing the range of ages at which sons are observed from 30-55 to 35-42 years old slightly increases the estimated IGE, thus eliminating some of the bias. Including an interaction term into the regression equation is another way to treat life-cycle bias.

Besides, the SOEP is a rather short survey. It started in 1984 and spans only three decades, which is too short to contain career-long information for two successive generations. A larger sample could enable to observe earnings for longer periods of time and at suitable ages for both generations, and thus to get more precise and reliable results. To further investigate the extent of intergenerational mobility in Germany, more waves of the SOEP should be used when available.

Chapter 2

Sibling correlations in France, compared to Sweden

2.1 Introduction

Recent public debates have echoed growing concern that, in modern democratic societies, a sizable share of economic inequality remains inherited within families. Assessing the role of family background in shaping individual economic success is indeed crucial to gauge the degree of inequality of opportunity in a society. Over the last twenty years, an important body of empirical research has investigated the extent of the intergenerational transmission of inequality (Solon, 1999; Björklund and Jäntti, 2009; Black and Devereux, 2011). This literature has demonstrated that between 20 and 60% of economic advantage is transmitted, within families, from one generation to the next. In the case of France, existing estimates indicate a value of the intergenerational earnings elasticity of about 0.5 (Lefranc, 2011). This suggests that about 25% of earnings inequality is transmitted within families across generations. However, by focusing on a single dimension of family characteristics, namely earnings, estimates of the intergenerational earnings elasticity fall short of fully accounting for the variety of channels through which the characteristics of the family influence the outcomes of their offspring. To provide a more comprehensive view, several recent papers have examined the degree of association in siblings' social and economic success, as surveyed in Björklund and Jäntti (2009, 2012) and Björklund and Salvanes (2010). Compared to the intergenerational correlation, the sibling correlation in socioeconomic outcomes allows to capture a broader set of family influences.

Sibling analysis is also employed in sociological studies to assess more than the impact of parental education or occupation, as in Boutchénik, Coron, Grobon, Goffette, and Vallet (2015), Sieben (2001) or Knigge (2015). Sibling correlations indeed provide a summary measure of all the effects attributed to factors shared by siblings (Björklund and Jäntti, 2009; Björklund, Lindahl, and Lindquist, 2010). It captures the overall impact of growing up in the same family, and thus allows a more complete investigation of the role of family background in inequalities. As explained in Solon (1999), Björklund and Jäntti (2009) and Bingley and Cappellari (2013), the sibling association in one particular measure of socioeconomic success (e.g. earnings, SES, education, etc.) can be decomposed as the sum of two terms. The first one is the square of the intergenerational correlation in the specific measure of socioeconomic success (e.g. the intergenerational correlation in earnings). The second one captures the influence of all the factors shared among siblings that are uncorrelated with the relevant parental measure of socioeconomic success.

Sibling correlations for various socioeconomic outcomes have been estimated in different countries, since early studies focusing on the US as Corcoran, Jencks, and Olneck (1976) or later Altonji and Dunn (1991), Corcoran et al. (1990) and Solon et al. (1991). In terms of education, sibling correlations are recently found to lie between 0.4 for Nordic countries (Björklund et al., 2009; Björklund and Salvanes, 2010; Björklund and Jäntti, 2012) and 0.6 for the United States (Conley and Glauber, 2008; Mazumder, 2008, 2011). Correlations in earnings lie between 0.2 for Nordic countries (Björklund et al., 2002; Björklund and Jäntti, 2009, 2012; Schnitzlein,

2014) and 0.4 for Germany (Schnitzlein, 2014) and the United States (Björklund et al., 2002; Conley and Glauber, 2008; Levine and Mazumder, 2007; Mazumder, 2008, 2011; Schnitzlein, 2014). Thus these sibling correlation estimates reveal a far more important transmission of inequalities than shown by estimated intergenerational elasticities. Nevertheless, if several authors have investigated the cases of other countries, the extent of sibling correlation in economic success has not been investigated for France. Boutchénik et al. (2015) study siblings' resemblance in terms of education and profession but not in terms of earnings. Our objective is to fill this gap in the literature.

Several features of France's socioeconomic setting make this country an interesting case for studying the role of family influences in shaping inequality of opportunity. Firstly, the French labor market is largely viewed as a heavily regulated one yielding a more compressed wage structure than observed in Anglo-Saxon economies. Secondly, deep reforms of the educational system have led to an important rise in access to higher education[1]. Furthermore, it is worth recalling that college, university and "grandes écoles" education are free of tuition in France. However, in international comparisons, France stands out as a country with low intergenerational mobility and high inequality of opportunity. In this context, it is worth investigating further the role of family background by relying on more comprehensive measure of family influences.

The objective of this paper is to measure the extent of sibling association in socioeconomic outcomes in France, and to compare our French results to sibling correlation in Sweden, both for methodological and comparative reasons. We use data from the French Education-Training-Employment (FQP) survey to estimate sibling similarities in different socioeconomic outcomes in France: profession, education and earnings. We find sibling correlations around 0.3, 0.4 and 0.5 for prestige

[1]Before 1975, lower secondary education was segmented into vocational and general schooling. This dual system was reformed in 1975 under the "réforme Haby" to create a unified junior high school curriculum. Access to higher education rose markedly, first in the late sixties-early seventies and again, during the late eighties-early nineties.

scores, annual earnings and education years respectively. When conducting a study by gender, it appears that same-sex siblings have more in common than in mixed pairs, for each outcome. Additional parameters are then investigated. They do not lead to any clear conclusion toward the evolution in time of sibling correlations. However concerning the impact of family composition, closely spaced siblings are more alike and family size seems to have a positive effect on sibling correlations. Finally we investigate the effect of parental education and profession but observe no clear pattern, except for the decrease of sibling correlations in earnings with higher educational levels of both parents.

We then construct comparable samples for France and Sweden and reestimate brother correlations in earnings for both countries. For Sweden, as earnings observations over career-long periods are available for siblings, we can estimate both correlations in permanent earnings and in predicted earnings, using the same method as for France. In addition to allowing us to compare the extent of intergenerational mobility in the two countries, it gives us an idea of the direction and extent of biases involved and enables us to better assess the actual correlation in permanent earnings in France. For Sweden, we estimate at 0.27 the brother correlation in permanent earnings – which is in line with the existing literature for this country – and around 0.35 the brother correlation in predicted earnings. In France, our estimation of the brother correlation in predicted earnings, around 0.54, suggests that the actual correlation in permanent earnings would lie around 0.4 for men.

The rest of the paper is organized as follows. Section 2.2 presents the estimation methods of sibling correlations. Section 2.3 presents the FQP data and describes the construction of the outcomes we further investigate. Section 2.4 reports the results obtained for France. Section 2.5 presents the specific model, data and results obtained in the comparison of France and Sweden. Section 2.6 concludes.

2.2 Methods

2.2.1 Pearson's and polychoric correlations

In order to estimate linear sibling correlation coefficients, let y_{ijt} be a continuous outcome of individual j in family i at time t. The importance of family background is measured by the share of variance of the long-run outcome that is accounted for by family effects. This "R^2" of family background is called the sibling correlation since it coincides with the correlation coefficient of randomly drawn pairs of siblings.

The modeled outcome y_{ijt} consists of a permanent part y_{ij} decomposed in a family component a_i, common to both siblings in family i and an individual-specific component b_{ij} for sibling j in family i, as well as a transitory error v_{ijt}:

$$y_{ijt} = y_{ij} + v_{ijt} = a_i + b_{ij} + v_{ijt}, \quad a \perp b \perp v. \tag{2.1}$$

The variance of the long-run outcome equals the sum of the variances of the family and individual components:

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2. \tag{2.2}$$

Thus the share of family background in the long-run outcome variance, the sibling correlation in permanent outcome ρ, is:

$$\rho = \frac{\sigma_a^2}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{2.3}$$

A set of complementary controls can be included in the estimation of the model in order to first purge the long-run outcome of some effects:

$$y_{ij} = \gamma Z_{ij} + a_i' + b_{ij}' = \gamma Z_{ij} + e_{ij}. \tag{2.4}$$

In particular the vector Z_{ij} contains in our French estimations a gender dummy, a quartic function of age, and all corresponding interactions. The residuals e_{ij} from the regression equation, free of gender and age effects, are then used in order to compute Pearson's correlations between two siblings 1 and 2:

$$\rho_{e_1,e_2} = \frac{cov(e_1, e_2)}{\sigma_{e_1}\sigma_{e_2}}. \tag{2.5}$$

In addition to linear sibling correlations, we also estimate polychoric correlations. Based on the same model, they measure the association between two ordinal variables assumed to be determined by a latent continuous variable following a bivariate normal distribution (Drasgow, 1988).

2.2.2 Sibling correlations linked to intergenerational elasticities

To link the sibling approach to the estimation of the transmission of socioeconomic outcomes using the intergenerational elasticity (IGE) β, we can further decompose our family component a_i into a part correlated to father's outcome X_i and all other family factors v_i uncorrelated with it:

$$a_i = \beta X_i + v_i. \tag{2.6}$$

Then son's outcome regressed on father's one is now:

$$y_{ij} = \beta X_i + u_{ij}, \text{ with } u_{ij} = v_i + b_{ij}. \tag{2.7}$$

Expressing the variance of the family component a_i yields:

$$\sigma_a^2 = \beta^2 \sigma_X^2 + \sigma_v^2, \tag{2.8}$$

and dividing both sides by σ_y^2, assuming same distributions among both generations, i.e. $\sigma_y = \sigma_X$, gives the sibling correlation ρ:

$$\rho = \frac{\sigma_a^2}{\sigma_y^2} = \beta^2 + \frac{\sigma_v^2}{\sigma_y^2}. \tag{2.9}$$

The sibling correlation ρ can thus be expressed as the sum of two terms: the squared of the IGE[2] β and a second component capturing the effect of all factors shared among siblings and uncorrelated with father's outcome. We can use this decomposition to consider how much of the sibling correlation is related to each part and thus how much is unaccounted for by the IGE.

2.2.3 Correlations on predicted variables

As further explained in Section 2.3, for the case of France, we do not estimate correlations on directly observed variables. Indeed extensive information is available for one of the siblings – hereafter "ego" – but only limited information is available for the other one – hereafter "alter". Instead we first predict continuous variables to then use them to investigate sibling correlations. We can model the latent outcome y as the sum of our predicted variable $\hat{y}$ and an ϵ, for each sibling:

$$y_j = \hat{y}_j + \epsilon_j, \text{with } j = E \text{ for ego, } j = A \text{ for alter.} \tag{2.10}$$

Considering that the distributions are the same for both siblings (that is $\sigma_{\hat{y_E}} = \sigma_{\hat{y_A}}$ and $\sigma_{\epsilon_E} = \sigma_{\epsilon_A}$) and that $\hat{y}$ and ϵ are independent (so $\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_\epsilon^2$), we find:

$$\rho(y_E, y_A) = \frac{cov(\hat{y}_E, \hat{y}_A) + cov(\epsilon_E, \epsilon_A)}{\sigma_{\hat{y}}^2 + \sigma_\epsilon^2} = \frac{\rho(\hat{y}_E, \hat{y}_A).\sigma_{\hat{y}}^2 + \rho(\epsilon_E, \epsilon_A).\sigma_\epsilon^2}{\sigma_{\hat{y}}^2 + \sigma_\epsilon^2}, \tag{2.11}$$

which means:

$$\rho(y_E, y_A) = \rho(\hat{y}_E, \hat{y}_A) \iff \rho(\hat{y}_E, \hat{y}_A) = \rho(\epsilon_E, \epsilon_A). \tag{2.12}$$

[2]If distributions of earnings are not assumed equal for both generations, the squared of the IGE is just replaced in equation (2.9) by the squared of the intergenerational correlation (IGC).

So if we assume that the sibling association in observed characteristics used to predict earnings is the same as in non observed characteristics, there is no impact of the use of predicted variables instead of observed ones, on the estimated sibling correlations. We further investigate the difference between sibling correlations in permanent and predicted earnings in Section 2.5, with the comparison of French and Swedish estimates.

2.3 Data

2.3.1 Description of the database and selection strategy

The data used in this paper come from the French Education-Training-Employment (FQP) survey. Targeted individuals are 18 to 65 year old people living in France, yielding a sample of around 40,000 individuals. We use as our main data set the wave of 2003, in which information on individuals and one of their siblings is available. Interviewed individuals (referred to as "ego") report personal information, notably on their education, occupation and earnings. They are also asked about their family environment and in particular the number of siblings, among whom one is randomly selected (referred to as "alter"). Individuals then also report some basic demographic and socioeconomic information about this sibling, among which education and occupation, but not earnings. The waves of 1970, 1977, 1985 and 1993 are also used as auxiliary data sets to help predicting continuous outcomes.

For our analysis we select individuals (ego) born between 1943 and 1973, which means 30 to 60 years old in 2003, to target individuals out of school but still in the labor market. We only keep individuals who reported information about a sibling. We allow up to 10 years of age difference between the individual (ego) and his/her sibling (alter). Therefore, siblings (alter) can be born between 1933 and 1983 and are 20 to 70 years old in 2003. This choice is made to avoid sampling young people with only older siblings, and old people with only younger siblings.

Available information concerning gender, birth cohort, education level and socio-professional category for both siblings – and earnings for ego – allow us to investigate sibling correlations in different socioeconomic outcomes. Additional information on the composition of the family – as number and birth order of brothers/sisters, age difference between ego and alter – and birth cohort, education, profession of the parents, enable taking various characteristics of the family into account to investigate their impact on sibling correlations.

In order to estimate linear sibling correlations, as seen in Section 2.2, we need continuous outcomes. In terms of education and occupation, continuous variables of years of education and prestige scores associated with the profession are constructed based on education level and socio-professional category. The predictions of these outcomes are respectively based on OLS regressions and scales of prestige scores. Earnings profiles are estimated to predict annual earnings at age 40 for both siblings using information on education and occupation of both ego and alter, and earnings of ego. OLS regressions as well as Heckman models are computed, to assess the issue of women's labor market participation.

2.3.2 Education

The available information for both siblings concerning education level is the highest completed certificate or degree. The corresponding variables in 8 groups (1 to 8 from highest degrees to no degree) are already ordinal ones, so that they are used to estimate polychoric correlation coefficients.

In order to predict as a continuous measure the number of years of education of individual i, $EducationYears_i$, for both siblings (this information being available only for ego, not for alter), we implement an OLS regression including as explanatory variables a dummy variable for the gender G_i, a quartic polynomial in age A_i, dummy variables Edu_i^k for the degree category k of individual i (with "no degree" as omitted category), and all corresponding interactions.

$$EducationYears_i = \alpha_1 G_i + \sum_{j=1}^{4} \alpha_{2j} A_i^j + \sum_{j=1}^{4} \alpha_{3j} G_i * A_i^j + \sum_{k=1}^{7} \alpha_{4k} Edu_i^k$$
$$+ \sum_{k=1}^{7} \alpha_{5k} Edu_i^k * G_i + \sum_{j=1}^{4} \sum_{k=1}^{7} \alpha_{6jk} Edu_i^k * A_i^j + \sum_{j=1}^{4} \sum_{k=1}^{7} \alpha_{7jk} Edu_i^k * G_i * A_i^j + u_i. \tag{2.13}$$

Waves of the survey of 1993 and 2003 are used in order to predict a number of education years for individuals born from 1933 to 1983. Indeed people can only be up to 65 years old in 2003 with the wave of 2003, and we need siblings "alter" aged up to 70 in 2003, which is why we also need the wave of 1993. The graphic representation for this regression is reported in Figure 2.1 for men (see Figure 2.3 in appendix for women).[3] This yields a prediction of 12.7 (std err. 2.84) and 13.0 (std err. 2.85) years of education on average for ego and alter respectively.

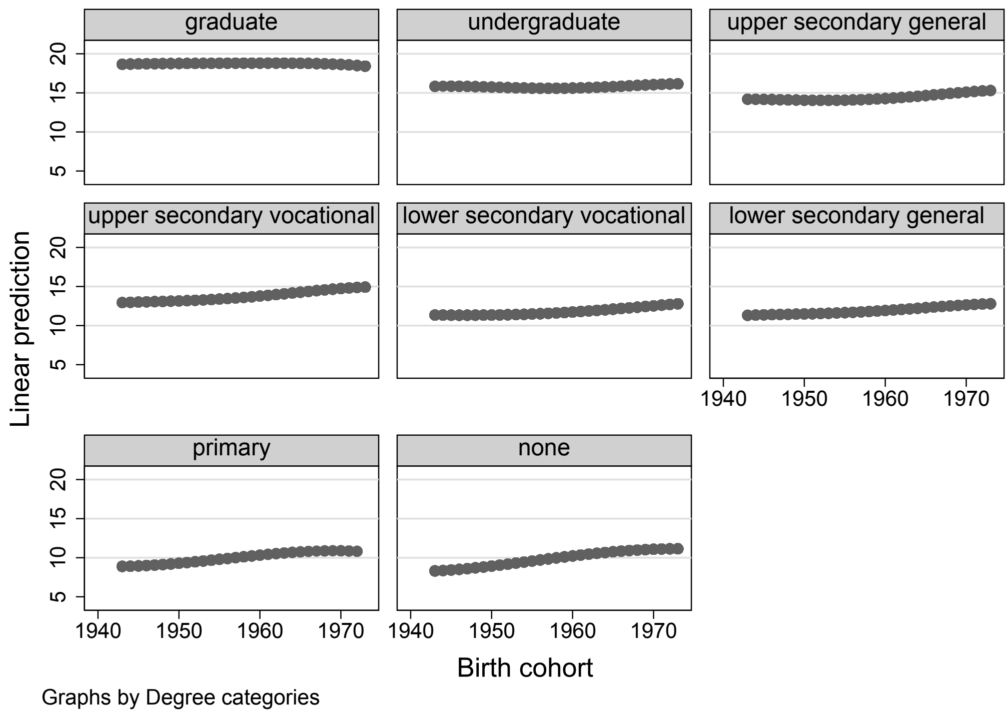


Figure 2.1: Predicted number of years of education for men

[3]A non-parametric specification including dummies for each gender/cohort/degree category is also tested and yields very similar results.

2.3.3 Occupation

Regarding occupation, a detailed classification containing 30 socio-professional groups is available in the data and used here. To construct an ordinal variable for the estimation of polychoric correlations, we gather some groups to obtain the following classification: 1) executive, manager, intellectual worker; 2) intermediate occupations; 3) skilled workman, craftspersons, storekeeper, company manager; 4) administrative, sales or service occupations and 5) farmers and laborers.

We also need a continuous measure of the occupation, and thus use prestige scores associated with professions, based on Chambaz, Maurin, and Torelli (1998). In their paper, they construct scales of prestige scores for different classifications of professions or socio-professional categories. We want to obtain a precise scale by attributing a score to each of our 30 groups. To do so we use the extremely detailed scale of scores attributed to a list of professions. The profession is however only available for ego in our data, so that we attribute the weighted mean of scores (weighted by the frequency of each profession in the groups) for each of our 30 groups of socio-professional categories, for both siblings. The distribution for this classification is reported in Table 2.15, in appendix.

2.3.4 Earnings

There is a measure of annual earnings in the wave 2003 of the survey, however only available for interviewed individuals (ego), not for their siblings (alter). The strategy to predict earnings for both siblings is here to estimate earnings profiles in a first step using all observable characteristics that are common to both siblings. We use all waves from 1970 to 2003 (1970, 1977, 1985, 1993 and 2003). Then in a second step we predict log of earnings for both siblings in the database of 2003.

We estimate earnings profiles based on individuals born between years 1933 and 1983 and observed from ages 25 to 55. We normalize age to zero at age 40, at which we predict earnings, in order to avoid life-cycle bias. We construct five groups of birth cohort covering 10 years each (1933-42, 1943-52, 1953-62, 1963-72 and 1973-83) as explanatory variables. The last two groups are actually reunited in the OLS estimation, because the last one contains individuals born from 1973, too late to predict a satisfactory earnings profile, and stops in fact in year 1978, no individual being younger than 25.

The dummy variables Edu_i^k corresponding to the different degree categories of education used in the construction of the continuous number of years of education are also used here as explanatory variables in the prediction of earnings (with again "no degree" as omitted category). Additionally the interactions of education levels with the dummy variables representing cohort groups Coh_i^l (with "1953-62" as omitted category), and with a quadratic function in age A_i, are included as regressors.

The ordinal classification of occupations is used as dummy variables Occ_i^m for interactions with cohort groups (with "laborers" as omitted category). We exclude categories "unknown","farmers" and "skilled workman, craftspersons, storekeeper, company manager", because most individuals of the two last ones are not salaried and therefore do not present a satisfactory measure of earnings. We thus restrict the sample to salaried individuals. We also use a more detailed classification of socio-professional categories as principal effects. This classification contains 16 categories used as dummy variables SPC_i^n (with "unskilled workers" as omitted category) of salaried individuals (only the clerical occupations are additionally excluded).

For men, the regression equation of the log of earnings y_{it} thus contains different age-earnings profiles based on education, as well as interactions between cohort groups and both education and occupation, and can be written:

$$\begin{aligned} y_{it} = \alpha_t &+ \sum_{j=1}^{2} \alpha_{1j} A_i^j + \sum_{l=1}^{3} \alpha_{2l} Coh_i^l + \sum_{n=1}^{15} \alpha_{3n} SPC_i^n + \sum_{m=1}^{3} \sum_{l=1}^{3} \alpha_{4ml} Occ_i^m * Coh_i^l \\ &+ \sum_{k=1}^{7} \alpha_{5k} Edu_i^k + \sum_{k=1}^{7} \sum_{l=1}^{3} \alpha_{6kl} Edu_i^k * Coh_i^l + \sum_{k=1}^{7} \sum_{j=1}^{2} \alpha_{7jk} Edu_i^k * A_i^j + u_{it}, \end{aligned} \tag{2.14}$$

where i and t are indices for individual and date of the survey. Then, we compute predicted log of earnings at age 40 in 2003 for both siblings.

To predict earnings for women, two alternative models are implemented: the same OLS regression as well as an Heckman model, in order to handle the issue of their participation into the labor force. Number of children and spouse's education level, contained in Z_i, are then additionally used to account for the probability of being in salaried employment, with y_{it} only observed for women when the following selection equation is satisfied:

$$f(age_i, Coh_i, Edu_i, Occ_i, SPC_i, Z_i) + v_{it} > 0. \tag{2.15}$$

To illustrate these earnings profiles, we represent earnings gains obtained for each level of education for the different birth cohort groups, as well as the effect of age on earnings also for each level of education and for the cohort group of reference, individuals born between 1953 and 1962. This is reported in Figure 2.2 for men (see Figure 2.4 in appendix for women).

2.3.5 Description of the sample

To obtain our final sample, we only keep individuals reporting education and occupation information for both siblings.[4] This results in a sample of 19,589 sibling pairs, among which 4,901 pairs of brothers and 4,732 pairs of sisters. The remaining 9,956 are mixed pairs. Ages and age differences among pairs of siblings are reported in Table 2.1. Siblings are aged 44 on average, with an average age difference of 4 years. We also report in this Table the average number of siblings in the family, which amounts 2.9. More precisely 5,489 family count two siblings only, 5,182 count three, 3,251 count four, the remaining 5,667 count five siblings or more.

[4]This strategy does not excessively reduce the sample size and allows to work with a more stable sample. From an initial sample of 21,885 pairs of siblings, 329 present missing occupation information for ego or alter and 2,212 regarding education.

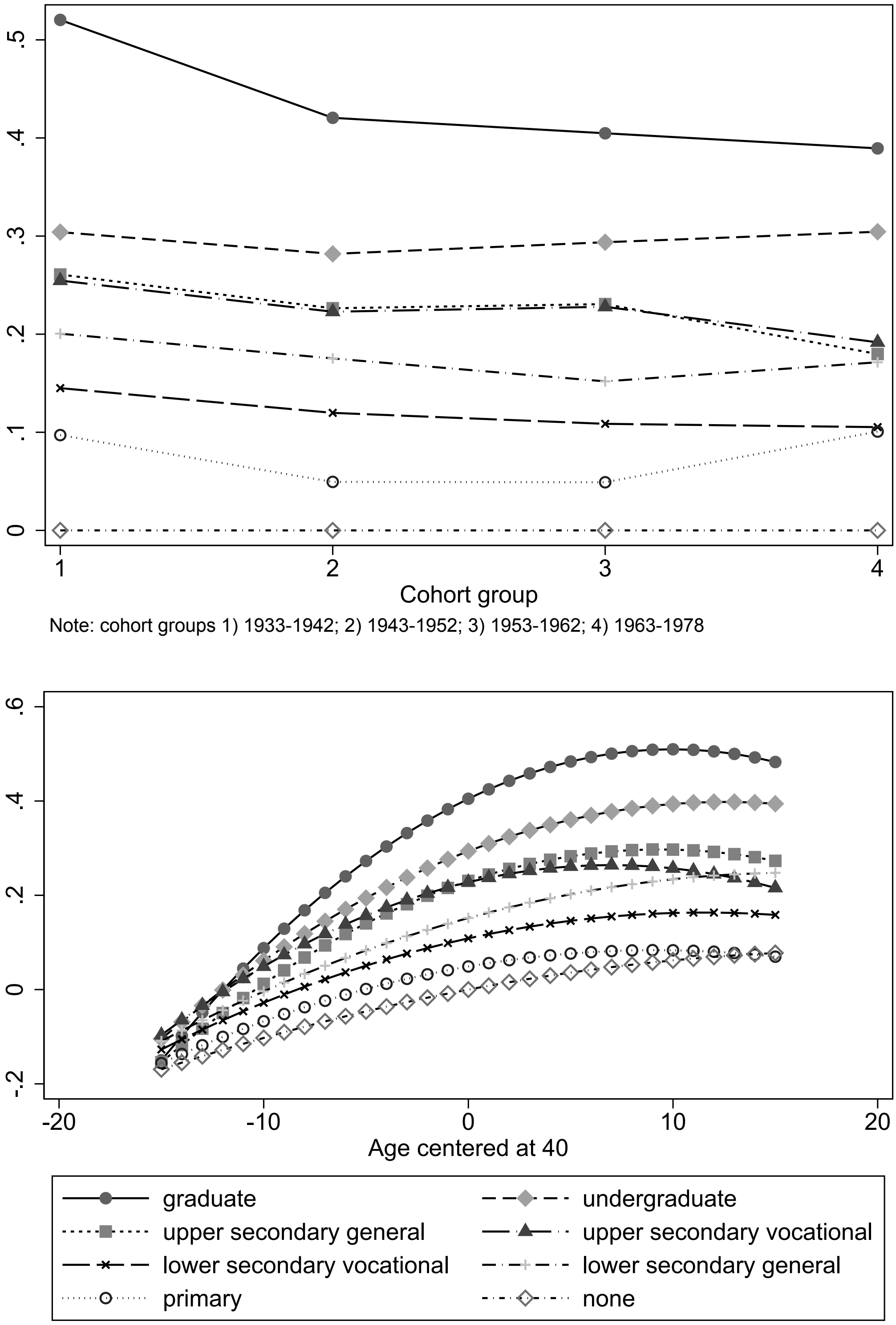


Figure 2.2: Earnings gains by education and cohort with "no degree" as reference, and returns to age by education for the group reference "born 1953-1962", for men

Table 2.1: Constitution of the families

	Mean	Std. Dev.	Min	Max	Obs.
Ego's age	44.254	8.603	30	60	19,589
Alter's age	44.267	9.653	20	70	19,589
Age diff.	4.234	2.555	0	10	19,589
0 to 3 years	2.047	0.811	0	3	9,248
4 to 6 years	4.869	0.813	4	6	6,348
7 to 10 years	8.291	1.127	7	10	3,993
Size of sibship	2.941	2.112	1	17	19,589

In Table 2.2 we present the distributions of the ordinal variables previously described representing degrees and socio-professional categories, for both siblings. For education as well as for occupation, distributions for ego and alter are close.

Table 2.2: Degrees and socio-professional categories

	Ego		Alter	
	Freq.	Percent	Freq.	Percent
Degree				
graduate	2,196	11.21	2,419	12.35
undergraduate	1,906	9.73	2,382	12.16
upper secondary general	1,302	6.65	1,764	9.01
upper secondary vocational	1,564	7.98	930	4.75
lower secondary vocational	5,411	27.62	6,402	32.68
lower secondary general	1,875	9.57	1,507	7.69
primary	1,682	8.59	1,592	8.13
none	3,653	18.65	2,593	13.24
Category				
executive, ...	2,786	14.22	1,996	10.19
intermediate occupations	4,736	24.18	4,431	22.62
skilled workman, ...	1,024	5.23	1,472	7.51
administrative, ...	6,032	30.79	5,993	30.59
farmer and laborers	5,011	25.58	5,697	29.08

Additionally we also compute ordinal variables representing education and occupation for the parents and the distributions are reported in Table 2.3. Regarding socio-professional categories, the variables are the same for both generations. For highest completed education, parents are aggregated in only three groups: at least upper secondary, lower secondary, and primary or none.

Table 2.3: Parental degrees and socio-professional categories

	Father		Mother	
	Freq.	Percent	Freq.	Percent
Degree				
upper secondary or more	2,695	14.13	2,007	10.31
lower secondary	3,739	19.61	2,873	14.75
primary or none	12,635	66.26	14,596	74.94
Category				
executive, ...	1,684	8.87	260	1.96
intermediate occupations	2,655	13.99	1,604	12.11
skilled workman, ...	2,372	12.49	1,215	9.17
administrative, ...	2,084	10.98	5,466	41.25
farmer and laborers	10,189	53.67	4,705	35.51

2.4 Results

2.4.1 Main results

We report in Table 2.4 estimates of polychoric correlations for education and occupation, as well as linear correlations for education, occupation and earnings, both directly using the continuous outcomes (referred to as "gross") and using residuals free of gender and age effects (referred to as "net").[5] The polychoric correlations amount respectively 0.553 and 0.375 for education and occupation, which is close to the gross linear estimates: 0.580 for education and 0.329 for occupation. The estimate is 0.446 for earnings which, when compared to the IGE estimated around 0.5 in Lefranc (2011) – thus corresponding to a sibling correlation of 0.25 if all family influences were accounted for through father's earnings – suggests that a substantial part of the effect of family background was in fact not captured.

These results are satisfactory since it was expected for sibling correlations to be higher in terms of education than occupation. Indeed education is likely to be more affected by family influences as it is determined at an earlier stage in life than occupation. Moreover annual earnings are here predicted based on both education

[5]Results obtained with different strategies to construct the continuous outcomes of education years and annual earnings lead to very similar results: 0.575 (0.005) for the non-parametric approach of predicting education and 0.410 (0.007) with OLS regression used for both gender instead of Heckman model for women, to predict earnings.

and occupation information, therefore it is also not surprising for sibling correlations in terms of earnings to lie in between.[6]

Table 2.4: Linear and polychoric sibling correlations

	Education		Occupation		Earnings
	Linear	Polychoric	Linear	Polychoric	Linear
gross	0.580	0.553	0.329	0.375	0.446
	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)
net	0.522		0.336		0.459
	(0.005)		(0.006)		(0.006)

Note: Linear corresponds to Pearson's correlations estimated on predicted variables (number of education years, prestige scores and ln of annual earnings at age 40 for education, occupation and earnings respectively), Polychoric corresponds to polychoric correlations estimated on ordinal variables (highest completed certificate or degree and socio-professional categories for education and occupation respectively).

These estimates are also in line with the international literature, as shown in the summary of some recent studies' results in various countries reported in Table 2.5. In terms of education as well as earnings, our estimates are higher than those of Nordic countries. For education they are smaller than those of the United States and for earnings they are close to those of the United States and Germany. This is coherent with the existing international ranking based on the estimation of intergenerational elasticities.

Whereas controlling for gender and cohort effects only slightly increases sibling correlations in terms of occupation and earnings, from 0.329 to 0.336 and from 0.446 to 0.459 respectively, it decreases the estimates for education from 0.580 to 0.522. An explanation is that education is more affected than occupation or earnings by the fact that siblings are often born in close cohorts, so that a general trend in the evolution of education level artificially raises the sibling correlation. We investigate net correlation coefficients from here on.

[6] The sample is reduced from 19,589 to 16,338 sibling pairs for the estimation of earnings correlations, thus we also estimate education and occupation correlations on this smaller sample: it only slightly increases the net estimates from 0.522 (0.005) to 0.533 (0.006) for education and from 0.336 (0.006) to 0.368 (0.007) for occupation (the evolution being similar on gross estimates).

Table 2.5: Recent estimates of sibling correlations in education and income

Country	Authors	Data	Cohorts/ages	Est.
Education				
Norway	Björklund and Salvanes (2010)	registers	1962-68	0.40 (0.01)
Sweden	Björklund et al. (2009)	registers	1962-68/30-38	0.48 (0.02)
	Björklund and Jäntti (2012)	registers	1951-67/≈40	0.44 (0.00)
United States	Conley and Glauber (2008)	PSID	1958-76/25-43	0.63 (0.07)
	Mazumder (2008)	NLSY	1957-64/26-41	0.62 (0.01)
	Mazumder (2011)	PSID	1951-68	0.67 (0.03)
Income				
Denmark	Björklund et al. (2002)	registers	1951-68/25-42	0.23 (0.01)
	Schnitzlein (2014)	registers	1952-76/30-50	0.20 (0.01)
Finland	Björklund et al. (2002)	registers	1953-65/25-42	0.26 (0.03)
Germany	Schnitzlein (2014)	SOEP	1952-78/30-50	0.43 (0.08)
Norway	Björklund et al. (2002)	registers	1950-70/25-42	0.14 (0.01)
Sweden	Björklund et al. (2002)	registers	1948-65/25-42	0.25 (0.01)
	Björklund et al. (2009)	registers	1962-68/30-38	0.37 (0.00)
	Björklund and Jäntti (2012)	registers	1951-67/31-40	0.22 (0.00)
United States	Björklund et al. (2002)	PSID	1951-67/25-42	0.43 (0.04)
	Conley and Glauber (2008)	PSID	1958-76/25-43	0.34 (0.07)
	Mazumder (2008)	NLSY	1957-65/26-41	0.49 (0.02)
	Mazumder (2011)	PSID	1951-68	0.51 (0.04)
	Levine and Mazumder (2007)	NLSY	1957-65/26-38	0.45 (0.05)
	Schnitzlein (2014)	PSID	1949-77/30-50	0.45 (0.04)

2.4.2 Sibling correlations by type of sibling pairs

Different sibling correlations are computed for same-sex (brother/brother and sister/sister) and mixed (brother/sister) sibling pairs and reported in Table 2.6. As expected, mixed sibling pairs share less than same-sex siblings for each outcome. Sisters seem to have a little more in common than brothers in terms of education and occupation (but these differences are not significant), and less regarding earnings: 0.467 for sisters and 0.517 for brothers.[7]

The relatively low participation of women into the labor force can raise an issue, since prestige scores are attributed according to the last reported socioeconomic category. Mostly for women, this can reflect the professional situation in the beginning of a short career, stopped for instance to raise children, whereas our interest is in obtainable prestige scores, potentially reached if everybody had always worked.

[7]See in appendix the method used for inference issues, based on Fisher (1915).

Table 2.6: Sibling correlations by gender

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
Brothers	0.543	0.352	0.517
	(0.013)	(0.014)	(0.014)
Sisters	0.551	0.377	0.467
	(0.011)	(0.013)	(0.012)
Mixed pairs	0.497	0.307	0.428
	(0.009)	(0.009)	(0.008)
p-values testing the equality of correlation coefficients			
Brothers/Sisters	0.589	0.159	0.003
Brothers/Mixed	0.000	0.003	0.000
Sisters/Mixed	0.000	0.000	0.010

A first solution to assess this issue is to only take into account currently working women. Therefore, we observe the restricted sample of women (ego) with a brother (alter). Sibling correlations between all women or only working women, and their brothers are reported in Table 2.7. They are presented for occupation as well as for education, to compare the effects on an outcome potentially affected by the employment of women and the other not. We also compare these results to the same obtained for men (ego) with brothers (alter).[8] As expected for education the results are almost not modified by sampling only currently working individuals as ego. But the differences are also small for occupation. And sampling according to the working status does not change the results more for women than for men.

However a selection problem can rise if the sample is restrained to currently working women. Another method is the investigation of brother/brother-in-law correlations. Again based on the sample of women with a brother, we construct prestige scores for women's spouses (socio-professional categories being available for them too), and we compare them to brothers' ones. The results are also reported in Table 2.7. The brother-in-law/brother correlation is not very different from, even if slightly lower than sister/brother correlations.

[8]Number of observations for the five groups in Table 2.7 are respectively 5,525, 4,420, 4,901, 4,527 and 5,525.

Table 2.7: Sibling correlations for all/working women

	Education	Occupation
All	0.522	0.336
	(0.005)	(0.006)
ego: woman; alter: man	0.494	0.303
	(0.012)	(0.012)
ego: working woman; alter: man	0.483	0.296
	(0.014)	(0.014)
ego: man; alter: man	0.543	0.352
	(0.013)	(0.014)
ego: working man; alter: man	0.551	0.358
	(0.011)	(0.015)
ego: husband; alter: man		0.272
		(0.015)

2.4.3 Effect of other characteristics on sibling correlations

We also take into account additional parameters, to investigate their impact on sibling correlations. First we want to investigate the evolution over the years of the effect of family background on siblings' outcomes. To do so, we split our sample into three groups, depending on the average parental birth cohort: before 1925, between 1925 and 1935, and after 1935, and estimate different sibling correlations for these three groups. We also test the same strategy based on average siblings' birth cohort: before 1954, between 1954 and 1964, and after 1964.[9] We report in Table 2.8 the results presenting the evolution of sibling correlations through time, however no clear pattern seems to be observed, so the correlation seems very stable over time.

Family and sibling pair characteristics are then considered, to investigate their effect on sibling correlations (Oettinger, 1999): age difference between ego and alter, number of siblings in the family and whether or not ego or alter is the oldest child of the sibship.[10] Estimates of sibling correlations obtained exploring these factors are reported in Table 2.9.

[9]Both sets of three groups – constructed based on parental and siblings' birth cohorts respectively – present a nearly perfect repartition in three thirds.

[10]In 12,027 sibling pairs ego or alter is the oldest child of the family, in the 7,562 others it is not the case.

Table 2.8: Evolution in time of sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
by parental birth cohort			
Before 1925	0.536	0.342	0.479
	(0.010)	(0.011)	(0.011)
1925-1935	0.512	0.346	0.457
	(0.010)	(0.013)	(0.011)
After 1935	0.514	0.321	0.443
	(0.010)	(0.010)	(0.010)
p-values testing the equality of correlation coefficients			
Before 1925/1925-1935	0.059	0.774	0.154
1925-1935/After 1935	0.843	0.096	0.341
Before 1925/After 1935	0.085	0.176	0.016
by siblings' birth cohort			
Before 1954	0.525	0.321	0.467
	(0.012)	(0.012)	(0.011)
1954-1964	0.514	0.351	0.456
	(0.011)	(0.010)	(0.010)
After 1964	0.526	0.333	0.456
	(0.010)	(0.012)	(0.013)
p-values testing the equality of correlation coefficients			
Before 1954/1954-1964	0.412	0.050	0.486
1954-1964/After 1964	0.344	0.255	0.978
Before 1954/After 1964	0.898	0.429	0.476

As expected, age difference has an impact on sibling correlations, at least when comparing closely spaced siblings to those with an important age gap: siblings seem to be more alike when they are about the same age. The estimates fall from 0.541 to 0.471 for education, from 0.347 to 0.312 for occupation and from 0.481 to 0.424 for earnings, for siblings with up to 3 years versus from 7 years age gap.

Concerning the effect of family size, correlations in education and earnings increase with the number of siblings, again the result being significant only when comparing families with substantial different sizes. The correlations increase for instance from 0.471 to 0.529 for education and from 0.410 to 0.440 for earnings, for families counting 2 versus at least 5 siblings.

Lastly the sibling correlation in terms of education is higher, 0.538 versus 0.498,

Table 2.9: Effect of family and sibling pair characteristics on sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
by age difference			
0 to 3 years	0.541	0.347	0.481
	(0.009)	(0.010)	(0.010)
4 to 6 years	0.523	0.334	0.450
	(0.012)	(0.013)	(0.013)
7 to 10 years	0.471	0.312	0.424
	(0.013)	(0.017)	(0.016)
p-values testing the equality of correlation coefficients			
0 to 3/4 to 6	0.123	0.395	0.030
4 to 6/7 to 10	0.001	0.207	0.138
0 to 3/7 to 10	0.000	0.038	0.001
by number of siblings			
2	0.471	0.308	0.410
	(0.012)	(0.014)	(0.015)
3	0.496	0.315	0.437
	(0.013)	(0.012)	(0.013)
4	0.510	0.303	0.438
	(0.013)	(0.015)	(0.014)
5 or more	0.529	0.313	0.440
	(0.013)	(0.013)	(0.013)
p-values testing the equality of correlation coefficients			
2/3	0.092	0.701	0.118
3/4	0.411	0.571	0.967
4/5 or more	0.240	0.627	0.909
2/4	0.021	0.813	0.159
3/5 or more	0.021	0.917	0.858
2/5 or more	0.000	0.774	0.076
whether one is the oldest child			
yes	0.498	0.331	0.441
	(0.008)	(0.009)	(0.008)
no	0.538	0.315	0.457
	(0.009)	(0.010)	(0.011)
p-values testing the equality of correlation coefficients			
yes/no	0.000	0.223	0.228

when neither ego nor alter is the oldest child of the family (Conley, 2009). This would indicate that the oldest child is more different from all other siblings than they are among each other, possibly because he or she is the only one who ever was an only child. The effect of family size may partly be driven by this last result, as it is more likely for either or alter to be the oldest child in smaller families (especially for sibships of only two siblings!).

Finally we want to observe the effect of parental characteristics, such as education and occupation, in order to further assess the impact family background can have on sibling correlations. Thus we report in Tables 2.10 and 2.11 the estimated correlation coefficients obtained for each educational level and socio-professional category of both parents. We also estimate these sibling correlations for the whole population, based on residuals net not only from siblings' age and gender effects, but also from education or socio-professional categories of the parents.

We can observe a decrease of sibling correlations in terms of education and earnings, with the increase of educational level of both parents. From lowest to highest completed education of the father, the estimates fall from 0.447 to 0.388 for education and from 0.406 to 0.303 for earnings; from lowest to highest completed education of the mother, they decrease from 0.450 to 0.387 for education and from 0.400 to 0.295 for earnings.

A possible explanation can lie in differences in investment strategies of reinforcement or compensation of sibling differences in initial endowments (Behrman, Pollak, and Taubman, 1982, 1986; Behrman and Taubman, 1986; Behrman, Rosenzweig, and Taubman, 1994) from more or less educated/wealthy parents. Indeed if parents care about the wealth of their children (more than about their earnings), the model of Becker and Tomes (1976) suggests that wealthy parents will invest the most efficient allocation in each child's human capital and then compensate any resulting earnings differences with financial transfers, whereas poorer parents only invest in their children's human capital, taking equality among their children as well as efficiency considerations into account. In this case, sibling differences in human capital and thus earnings are likely to increase with family wealth and education, as we observe.

Concerning the effect of parental occupation, sibling correlations often seem to be lower when parents' socio-professional categories are the highest: *executive, manager, intellectual worker*, which is coherent with the previous interpretation. No other clear pattern is observable.

Table 2.10: Effect of parental education on sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006
Father – net also from father's education	0.419	0.260	0.366
	(0.006)	(0.007)	(0.007)
1) upper secondary or more	0.388	0.248	0.303
	(0.020)	(0.020)	(0.019)
2) lower secondary	0.412	0.232	0.342
	(0.015)	(0.014)	(0.017)
3) primary or none	0.447	0.282	0.406
	(0.007)	(0.008)	(0.008)
p-values testing the equality of correlation coefficients			
1/2	0.271	0.490	0.104
2/3	0.020	0.004	0.000
1/3	0.001	0.083	0.000
Mother – net also from mother's education	0.431	0.267	0.375
	(0.006)	(0.007)	(0.007)
1) upper secondary or more	0.387	0.224	0.295
	(0.023)	(0.021)	(0.021)
2) lower secondary	0.423	0.257	0.351
	(0.017)	(0.016)	(0.016)
3) primary or none	0.450	0.281	0.400
	(0.008)	(0.007)	(0.009)
p-values testing the equality of correlation coefficients			
1/2	0.135	0.223	0.043
2/3	0.103	0.209	0.009
1/3	0.001	0.010	0.000

Table 2.11: Effect of parental occupation on sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
Father – net also from father's occupation	0.429	0.251	0.360
	(0.006)	(0.007)	(0.007)
1) executive, ...	0.380	0.210	0.303
	(0.020)	(0.022)	(0.023)
2) intermediate occupations	0.447	0.266	0.335
	(0.016)	(0.018)	(0.017)
3) skilled workman, ...	0.452	0.257	0.392
	(0.017)	(0.018)	(0.021)
4) administrative, ...	0.435	0.276	0.384
	(0.018)	(0.019)	(0.018)
5) farmer or laborer	0.428	0.249	0.369
	(0.009)	(0.009)	(0.010)
p-values testing the equality of correlation coefficients			
1/2	0.009	0.055	0.265
2/3	0.840	0.719	0.039
3/4	0.491	0.480	0.783
4/5	0.712	0.221	0.498
1/3	0.006	0.119	0.004
2/4	0.609	0.706	0.069
3/5	0.195	0.718	0.316
1/4	0.042	0.030	0.007
2/5	0.273	0.398	0.096
1/5	0.028	0.115	0.007
Mother – net also from mother's occupation	0.441	0.263	0.376
	(0.007)	(0.008)	(0.009)
1) executive, ...	0.367	0.229	0.229
	(0.057)	(0.058)	(0.057)
2) intermediate occupations	0.445	0.245	0.344
	(0.022)	(0.025)	(0.022)
3) skilled workman, ...	0.428	0.225	0.384
	(0.300)	(0.029)	(0.030)
4) administrative, ...	0.449	0.286	0.389
	(0.011)	(0.012)	(0.013)
5) farmer or laborer	0.440	0.252	0.381
	(0.014)	(0.015)	(0.016)
p-values testing the equality of correlation coefficients			
1/2	0.166	0.801	0.077
2/3	0.577	0.573	0.286
3/4	0.409	0.040	0.855
4/5	0.580	0.068	0.656
1/3	0.296	0.948	0.021
2/4	0.861	0.123	0.081
3/5	0.637	0.367	0.933
1/4	0.125	0.341	0.008
2/5	0.835	0.793	0.175
1/5	0.174	0.702	0.014

2.5 France-Sweden comparison

So far, we estimated sibling correlations based on predicted earnings for France. However we suspected that the estimation could be biased by the use of predicted measures instead of actual permanent earnings, as mentioned in Section 2.2.3. Therefore, we now compare French results to estimated brother correlations for Sweden, where extensive available information allow us to assess this bias.

2.5.1 Methods

Ideally we want to estimate the sibling correlation in permanent earnings. For the case of Sweden, we have access to multiple years of income observations for members of the same family, which enable us to directly estimate the actual correlation ρ for two randomly drawn siblings, ego (E) and alter (A):

$$\rho = \frac{\mathrm{Cov}[y_{iE}, y_{iA}]}{\sigma_{y_E}\sigma_{y_A}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{2.16}$$

In order to assess the sensitivity of the correlation to transitory characteristics, we also consider the sibling correlation in current instead of permanent income. Since the variance of current income includes the variance of the transitory error in addition to the sum of the variances of the family and individual components, the sibling correlation is then underestimated:

$$\rho_t = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma_v^2} < \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} = \rho. \tag{2.17}$$

If annual earnings for pairs of siblings are not available over career-long periods, as in France, we must resort to alternative methods of estimation by first predicting earnings, as we saw. Then we can estimate two alternative sibling correlations: $\tilde{\rho}$, the covariance between ego's and alter's predicted earnings, divided by the variance of permanent earnings; and $\tilde{\tilde{\rho}}$, the same covariance between ego's and alter's predicted earnings, divided by the variance of predicted earnings, i.e. the correlation coefficient in predicted earnings we estimated so far:

$$\tilde{\rho} = \frac{\mathrm{Cov}[\hat{y}_{iE}, \hat{y}_{iA}]}{\sigma_y^2}, \tag{2.18}$$

$$\tilde{\tilde{\rho}} = \frac{\mathrm{Cov}[\hat{y}_{iE}, \hat{y}_{iA}]}{\sigma_{\hat{y}}^2}. \tag{2.19}$$

To predict earnings for pairs of siblings, suppose we have K (time invariant) predictors of earnings, each consisting of a family and an individual component:

$$\tilde{x}_{ij,k} = x_{i,k} + x_{ij,k} \quad \text{with variances} \quad \sigma_{i,k}^2 + \sigma_{ij,k}^2. \tag{2.20}$$

Income can then be expressed as the sum of a part explained by the observed predictors and a part explained by unobserved characteristics:

$$y_{ijt} = \beta_0 + \sum_{k=1}^{K} \beta_k (x_{i,k} + x_{ij,k}) + e_i + e_{ij} + u_{ijt}. \tag{2.21}$$

The family and individual components of equation (2.1) can also be decomposed into these two parts, explained or not by the predictors:

$$a_i = \sum_{k=1}^{K} \beta_k x_{i,k} + e_i, \tag{2.22}$$

$$b_{ij} = \sum_{k=1}^{K} \beta_k x_{ij,k} + e_{ij}. \tag{2.23}$$

The variance of long-run income is then the weighted sum of the individual and family components' variances of the predictors and the residual parts (ignoring covariance across different predictors):

$$\sigma_y^2 = \sum_{k=1}^{K} \beta_k^2 (\sigma_{i,k}^2 + \sigma_{ij,k}^2) + \sigma_{e_i}^2 + \sigma_{e_{ij}}^2. \tag{2.24}$$

Similarly, the covariance of two siblings' income, i.e. σ_a^2 the variance of the family component, is:

$$\text{Cov}[y_{iEt}, y_{iAs}] = \sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2 + \sigma_{e_i}^2. \tag{2.25}$$

So we can think of the fraction:

$$\frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2}{\sigma_a^2} := \varrho_C^2 \tag{2.26}$$

as the fraction of the variance of the family component which is captured by the predictors, in the population. In the same way, we can express the population value of the fraction of the total variance of income captured by the predictors as:

$$\frac{\sum_{k=1}^{K} \beta_k^2 (\sigma_{i,k}^2 + \sigma_{ij,k}^2)}{\sigma_y^2} := \varrho^2. \tag{2.27}$$

This allows us to express the two alternative sibling correlations $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$ of equations (2.18) and (2.19) based on ϱ_C^2 and ϱ^2 – the parameters representing the fractions of family and total variances explained by the predictors – and to link them to the sibling correlation in permanent earnings, ρ:

$$\begin{aligned} \tilde{\rho} &= \frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2}{\sigma_y^2} = \frac{\varrho_C^2 \sigma_a^2}{\sigma_y^2} = \varrho_C^2 \rho, && (2.28) \\ \tilde{\tilde{\rho}} &= \frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2}{\sum_{k=1}^{K} \beta_k^2 (\sigma_{i,k}^2 + \sigma_{ij,k}^2)} = \frac{\varrho_C^2 \sigma_a^2}{\varrho^2 \sigma_y^2} = \frac{\varrho_C^2}{\varrho^2} \rho. && (2.29) \end{aligned}$$

Since $\varrho_C^2, \varrho^2 \in (0, 1)$, we know that $\tilde{\rho} - \rho < 0$ but we cannot, a priori, determine the sign of the difference $\tilde{\tilde{\rho}} - \rho$. Thus, $\tilde{\rho}$ underestimates the sibling correlation and provides a lower bound estimate. Indeed the covariance of predicted earnings is lower than the covariance of permanent earnings, since the variance of unobserved factors is unaccounted for:

$$\tilde{\rho} = \frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2}{\sigma_y^2} < \frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2 + \sigma_{e_i}^2}{\sigma_y^2} = \rho. \tag{2.30}$$

Regarding the sibling correlation in predicted earnings, $\tilde{\tilde{\rho}}$, the covariance of predicted earnings – the numerator – is still lower than the covariance of permanent earnings. However the variance of predicted earnings – the denominator – is also lower than the variance of permanent earnings:

$$\tilde{\tilde{\rho}} = \frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2}{\sum_{k=1}^{K} \beta_k^2 (\sigma_{i,k}^2 + \sigma_{ij,k}^2)} \quad \text{vs} \quad \frac{\sum_{k=1}^{K} \beta_k^2 \sigma_{i,k}^2 + \sigma_{e_i}^2}{\sum_{k=1}^{K} \beta_k^2 (\sigma_{i,k}^2 + \sigma_{ij,k}^2) + \sigma_{e_i}^2 + \sigma_{e_ij}^2} = \rho. \tag{2.31}$$

Thus whether $\tilde{\tilde{\rho}}$ under- or overestimates ρ depends on the predictors. In particular, the relationship between ρ and $\tilde{\tilde{\rho}}$ can be written as:

$$\rho = \varrho^2 \tilde{\tilde{\rho}} + (1 - \varrho^2) \frac{\sigma_{e_i}^2}{\sigma_{e_i}^2 + \sigma_{e_{ij}}^2}. \tag{2.32}$$

Hence $\tilde{\tilde{\rho}} = \rho$ if and only if $\frac{\sigma_{e_i}^2}{\sigma_{e_i}^2 + \sigma_{e_{ij}}^2} = \rho$. Whether this condition is verified or not depends on the predictors $\tilde{x}_{ij,k}$, since it would mean for the correlation in unobserved characteristics to be the same among siblings as the correlation in observed predictors, as discussed in Section 2.2.3. For instance, if the predictors only include common family characteristics $\frac{\sigma_{e_i}^2}{\sigma_{e_i}^2 + \sigma_{e_{ij}}^2} < \rho$ and $\tilde{\tilde{\rho}} > \rho$.

For Sweden, we are able to compare the actual correlation in permanent earnings, ρ, to several correlations in predicted earnings, $\tilde{\tilde{\rho}}$, obtained from different sets of predictors. Thus we can assess the relations between ρ and the different $\tilde{\tilde{\rho}}$ and, if these relations are assumed to be the same in Sweden and in France, apply them to the French estimations of $\tilde{\tilde{\rho}}$ to shed light on the actual value of ρ in France. This would mean to assume that the ratio $\frac{\tilde{\tilde{\rho}}}{\rho}$ is the same in the two countries, i.e. as:

$$\frac{\tilde{\tilde{\rho}}}{\rho} = \frac{\varrho_C^2}{\varrho^2}, \tag{2.33}$$

that the relation between the fractions of σ_a^2 and σ_y^2 captured by the predictors is the same. This last assumption depends on the relative roles played by the predictors in the determination of (the family component of) income, in France and in Sweden.

2.5.2 Data

In order to compute comparable correlation estimates for France and Sweden, we construct new data sets, restricted to men. For Sweden we use registers data from 1968 to 2007, which allows us to estimate the permanent and transitory parts of earnings, as well as to estimate sibling correlations of both permanent and predicted earnings. For France, we use the waves 1977, 1985, 1993 and 2003 of the FQP survey to estimate sibling correlations based on predicted earnings. We also use data from the Annual Declarations of Social Data (DADS), a data set constructed from the annual declaration of all French firms to the fiscal administration of the total earnings paid to each of their employees. We use the earnings panel over the period 1976-2010 (and restrict it to the same birth cohorts as in the main sample) for the decomposition of earnings into permanent and transitory parts.

For Sweden we first compute permanent earnings as the mean of all annual earnings available over the selected period (all years from 1968 to 2007), for men with at least 20 observations available (which excludes only 1.5% of the sibling pairs). For both countries, we also predict earnings at age 40 for both siblings, based on men born between 1933 and 1977 and aged 20 to 64. For comparability issues, we only use annual earnings observations for years available in France (1977, 1985, 1993 and 2003), for the prediction of earnings in France and Sweden. We use different specifications based on various sets of predictors, including notably education, occupation and/or size of the sibship.

To estimate sibling correlations, we then further restrict both our samples to brothers with 10 years of age difference at most and with ego born between 1943 and 1967 and alter between 1933 and 1977. Since the variable of interest is earnings, the analysis also excludes self-employed individuals. As reported in Table 2.12, regarding samples used to estimate brother correlations in predicted earnings, we observe 79,600 brother pairs for Sweden, with on average individuals aged 47, with 4.1 years of age difference between ego and alter and 2.6 siblings in the family. For

France, the sample consists of 3,009 brother pairs, with on average individuals aged 46, with 4.3 years of age difference between ego and alter and 4.2 siblings in the family.

Table 2.12: Descriptive statistics of the samples used for the estimation of brother correlations in predicted earnings for France and Sweden

	France				**Sweden**			
	Ego		Alter		Ego		Alter	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Age	46.52	(6.89)	46.22	(8.11)	47.50	(7.27)	47.43	(8.32)
Age gap	4.27	(2.57)	-	-	4.11	(2.25)	-	-
Sibsize	4.16	(2.19)	-	-	2.58	(0.85)	-	-

Source: Registers data for Sweden, FQP data for France, for years 1977, 1985, 1993 and 2003.
Sample: Salaried brothers with ego born 1943-1967 and alter 1933-1977, with maximum 10 years age gap.

2.5.3 Results

Sweden - a comparison of methods

For Sweden, we estimate long-run income for both siblings as the average of annual earnings – over 36 years on average – from individuals with at least 20 available earnings observations. The sibling correlation ρ is then estimated at 0.265, from a sample of 265,256 sibling pairs.

In order to assess the sensitivity of the estimation to transitory characteristics, we also estimate the sibling correlation in current earnings of a single year, here 2003. The sibling correlation uncorrected from transitory effects is underestimated as expected and estimated at 0.125, from a sample of 225,497 sibling pairs. The variance of permanent earnings was estimated at 0.325 in the previous estimation and the variance of current earnings is here estimated at 0.597,[11] suggesting that slightly more than half of the variance of current earnings is due to the permanent part, in Sweden.

We additionally compute a decomposition of individual earnings dynamics into permanent and transitory components. We suppose uncorrelated transitory compo-

[11] These variances are for alter, the corresponding ones for ego are 0.316 and 0.624 respectively.

nents, following a first order autoregressive process (AR(1)) or a first order moving average process (MA(1)), alternatively. The sample of 6,822,922 earnings observations from 342,027 individuals (20 observations per individual on average) presents a total variance of earnings of 0.709. The three specifications lead to an estimated variance of the permanent component at respectively 0.526, 0.391 and 0.546. The specification assuming the transitory component to follow an AR(1) process seems to be the closest to the results found from estimated current and permanent earnings, with a ratio of permanent to total variance of 55% (74% and 77% for the uncorrelated and MA(1) specifications respectively), as reported in Table 2.13.

Table 2.13: Variance decomposition of earnings for Sweden and France

	Transitory component: Uncorr. (1)	AR(1) (2)	MA(1) (3)	Estimated variances (4)
Sweden				
Total (T)	0.709	0.709	0.709	0.597
Permanent (P)	0.526	0.391	0.546	0.325
Ratio (P/T)	0.742	0.551	0.770	0.545
France				
Total (T)	0.566	0.566	0.566	-
Permanent (P)	0.391	0.362	0.391	-
Ratio (P/T)	0.690	0.639	0.690	-

Source: Registers data of 1968-2007 for Sweden, DADS data of 1976-2010 for France.
Sample: Salaried men born 1933-1977, for (1), (2) and (3) aged 30-55, with at least 5 earnings observations available, for (4) aged 20-64, with current earnings of 2003 and the average of all (min. 20) available earnings observations as permanent earnings.

We then want to estimate the two alternative sibling correlations, $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$, based on predicted earnings. To assess the sensitivity of correlation estimates to the set of earnings predictors, we perform variations in the regression equation. Additionally to year dummies, a quadratic function of age and a cohort group effect, we alternatively use either or both education level and/or occupation, interacted with the cohort groups – and with the age effect for education – using the same specification as presented in Section 2.3.4 for France. Then we also add a common family characteristic to education and occupation: the size of the sibship, interacted with the cohort groups. Estimates are reported in Table 2.14.

Table 2.14: Estimation of the brother correlations in permanent, current and predicted earnings for Sweden and predicted earnings for France

	Permanent earnings		Current earnings (2003)	
Sweden				
ρ	0.265		0.125	
Cov	0.085		0.076	
N	265,256		225,497	
	Predicted earnings:			
	Education*	Occupation*	Edu+Occ*	Edu+Occ+Sibsize*
Sweden				
$\tilde{\tilde{\rho}}$	0.425	0.242	0.347	0.349
$\tilde{\rho}$	0.049	0.033	0.077	0.077
Cov	0.016	0.011	0.025	0.025
R^2	(0.15)	(0.17)	(0.20)	(0.20)
N	79,600	79,600	79,600	79,600
France				
$\tilde{\tilde{\rho}}$	0.591	0.402	0.536	0.544
$\tilde{\rho}$	0.135	0.086	0.155	0.157
Cov	0.049	0.031	0.056	0.057
R^2	(0.25)	(0.34)	(0.36)	(0.36)
N	3,009	3,009	3,009	3,009

* Prediction of earnings based only on education, only on occupation, on education and occupation, or on education, occupation and size of the sibship.

Source: Registers data of 1968-2007 for Sweden, FQP data of 1977, 1985, 1993 and 2003 for France.

Sample: Salaried brothers aged 20-64, with ego born 1943-1967 and alter 1933-1977, with maximum 10 years age gap.

As mentioned, whether estimates of $\tilde{\tilde{\rho}}$ under- or overestimate ρ depends on the set of predictors. Education is more correlated among siblings than earnings, whereas occupation appears slightly less correlated than earnings, thus the correlation in earnings is over- (resp under-)estimated when only education (resp. occupation) is used to predict earnings: 0.425 (resp. 0.242) instead of 0.265. As the correlation in occupation – and thus the correlation in earnings predicted based on occupation – is closer to the correlation in permanent earnings, when both occupation and education are included in the prediction, the correlation in earnings is still overestimated, at

0.347. Including common family characteristics to the prediction of earnings leads to further overestimation of the correlation in earnings. When the number of siblings in the family is additionally used, the estimate amounts 0.349. Overall, the estimate obtained from the prediction of earnings based only on occupation seems the most accurate one and is close to the actual correlation in permanent earnings.

We also estimate the sibling correlation $\tilde{\rho}$ by dividing the covariance between ego's and alter's predicted earnings (for each regression specification) by the variance of permanent earnings, 0.325. As discussed previously, $\tilde{\rho}$ underestimates ρ for all specifications, since the covariance of predicted earnings is lower than the covariance in permanent earnings. The former grows closer to the latter as more predictors are included for the prediction. The estimate is 0.016 (resp. 0.011) when earnings are predicted only from education (resp. occupation), and 0.025 when both are used, and when the size of the sibship is further included. All estimated covariances of predicted earnings are still far from the covariance of permanent earnings, 0.085. As much individual as well as common family characteristics should be used to underestimate the covariance as little as possible.

These results suggest an estimation of the parameter ϱ_C^2 – the fraction of the variance of the family component, σ_a^2, captured by the predictors – between 0.12 and 0.29, and an estimation of the parameter ϱ^2 – the fraction of the total variance, σ_y^2, captured by the predictors – between 0.12 and 0.22.

France and Sweden compared

For the case of France, we cannot estimate the sibling correlation in permanent earnings, as annual earnings over career-long periods for pairs of siblings are not available, as addressed. Nonetheless we can estimate $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$ and compare the results to those obtained for Sweden. First, it sheds light on the relative extent of sibling correlation in earnings in the two countries. Second, it enables us to assess the actual sibling correlation in permanent earnings in France.

In order to assess the variance of permanent earnings – to estimate $\tilde{\rho}$ – we compute the same variance decompositions as for Sweden, with the French earnings panel. The sample of 879,797 earnings observations from 64,727 individuals (14 observations per individual on average) presents a total variance of earnings of 0.566. The three specifications lead to an estimated variance of permanent earnings at 0.391, 0.362 and 0.391 respectively. The corresponding ratio of permanent to total variance are 69%, 64% and 69% respectively, as reported in Table 2.13.

Thus the variance of current earnings is estimated smaller in France (0.566) than in Sweden, when the variance decomposition is used (0.709). However, the French estimate is close to the Swedish one, when the latter is obtained from estimated current and permanent earnings (0.597). The variance of permanent earnings also appears smaller in France than in Sweden, at least slightly as in the specification assuming transitory errors to follow an AR(1) process. Regarding the ratio of permanent to total variance of earnings, it depends on the specification. It seems more stable in France than in Sweden, and thus is smaller in France than in Sweden when uncorrelated or MA(1) transitory components are considered, and higher when assuming transitory components following an AR(1) process.

We then predict earnings at age 40 for both siblings and use these predicted earnings to estimate the brother correlations $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$. The results are presented is Table 2.14. The estimates of $\tilde{\rho}$ are again substantially smaller than the estimates of $\tilde{\tilde{\rho}}$, and range from 0.086 to 0.157 depending on the specification. They increase with the number of predictor included, as the covariance also increases.

The brother correlation in predicted earnings $\tilde{\tilde{\rho}}$ is estimated at 0.591 when only education is used to predict earnings, at 0.402 when only occupation is used, at 0.536 with both of them and at 0.544 when the size of the sibship is further included. Again the estimates are higher when education is used as predictor and when a family characteristic is included, and smaller when earnings are predicted based on occupation. As noted above, we estimated the French brother correlation in predicted earnings at 0.517 in Section 2.4.2, with a prediction of earnings based on education and occupation. So the results found here are coherent with the ones presented so far.

Overall estimates of $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$ are significantly higher for France than for Sweden, confirming more association between siblings' earnings and thus less economic mobility in France than in Sweden. Estimates of $\tilde{\tilde{\rho}}$ are in particular 1.4 to 1.7 times higher in France than in Sweden (estimates of $\tilde{\rho}$ are 2 to 2.8 times higher in France than in Sweden, but as mentioned seem to be less cogent estimates of the actual sibling correlation in permanent earnings, ρ).

If the ratios between the brother correlations in permanent and predicted earnings are assumed to be the same in Sweden and in France – and we multiply the French estimates of $\tilde{\tilde{\rho}}$ by the ratios $\frac{\rho}{\tilde{\tilde{\rho}}}$ obtained from the Swedish estimations – the results for each specification suggest a brother correlation in permanent earnings at 0.369, 0.440, 0.409 and 0.413 respectively, for France. If we further use the ratio obtained when Swedish earnings are predicted based on education and occupation to correct the estimated sibling correlation we obtained for all types of sibling pairs in Section 2.4, 0.459, this suggests a value of 0.351. However the sibling correlation estimated for all types of sibling pairs was smaller than for brothers and the correction might not be fitted for sisters and mixed pairs.

Overall, as there is no evidence supporting the assumption made, these conjectures should be considered with caution. It still seems reasonable to conclude that the actual brother correlation in permanent earnings in France is around 0.4 and the one for all types of sibling pairs slightly smaller. Regarding the R^2 of the prediction equations, education seems to capture a larger part of what can be explained by the predictors in Sweden than in France, whereas the opposite is true for occupation.

2.6 Conclusion

This paper investigates intergenerational mobility in France through sibling correlations, using data from the French Education-Training-Employment (FQP) survey. We study the impact of family background on different socioeconomic outcomes of adult children – education, occupation and earnings – in order to assess the share of inequalities due to family environment.

First, for two siblings in each family we construct ordinal outcomes of degrees and socio-professional categories, and predict continuous numbers of education years, prestige scores associated with the profession and annual earnings. We then compute polychoric and linear sibling correlations. In the main analysis, we find estimated correlations of 0.522 for education, 0.336 for occupation and 0.459 for earnings.

We also measure the effect of some personal and family characteristics on these sibling correlations. The most significant result is that same-sex sibling pairs share more similarities than mixed pairs. We find that family composition also has an impact. For instance sibling correlations increase with the number of siblings in the family. Finally parental education and socio-professional levels tend to decrease sibling correlations.

Additionally, we compute comparable results for France and Sweden in order, on the one hand to assess the validity of the method used for French data and thus the actual extent of sibling correlations in permanent earnings, on the other hand to compare the level of mobility in these two countries. Restricting the analysis to brothers, for France we estimate correlations between 0.402 and 0.591 from predicted earnings, in line with our previous results. For Sweden, these estimates range from 0.242 to 0.425 depending on the specification used to predict earnings, and we find a brother correlation in permanent earnings at 0.265. This suggests that the French brother correlation in permanent earnings would lie around 0.4.

Our results allow to compare the situation in France with the recent international literature on sibling correlations. In terms of education, results are a bit higher than 0.4 for Nordic countries, which present a high mobility, and 0.6 for the United States, at the other end of the scale. It is not surprising for our results to lie in between. Concerning earnings, our estimates are close to the German ones. Indeed for Germany sibling correlations in terms of income amount around 0.4 as ours, slightly lower than American ones and higher than the estimates around 0.2 for Nordic countries.

Furthermore our estimated sibling correlations bring a new perspective on the importance of inequality transmission in France, so far investigated with intergenerational elasticities. Indeed the sibling correlation can be expressed as the sum of the squared IGE on the one hand, and the other shared factors uncorrelated to father's earnings on the other hand, as mentioned. Thus if we consider an IGE estimated around 0.5 in Lefranc (2011), which would correspond to a sibling correlation of 0.25 if all family influences were captured through father's earnings, it seems that a large part – around 30% – of the transmission had not been accounted for. So the transmission of inequalities is more important than previously estimated and factors shared by siblings unrelated to parental income play a major role in it.

Thus by presenting sibling correlations for different socioeconomic outcomes, as well as the impact some family characteristics can have on them, this paper constitutes a first step to fill the gap in the literature on sibling correlations in France. It updates the amount and constitution of the French inequality transmission, and confirms the rank of France on this matter between Nordic countries and the United States, and close to other Western European countries.

2.A Appendix

Prediction of continuous outcomes

Education years

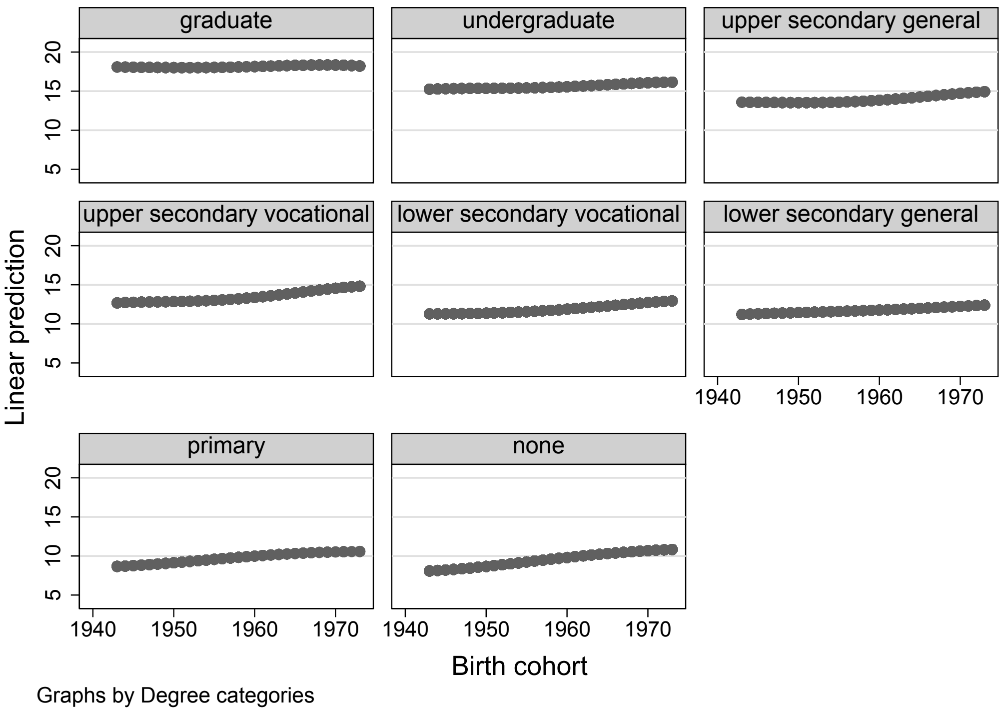


Figure 2.3: Predicted number of years of education for women

Prestige scores

Table 2.15: Prestige score – 30 groups

	Ego		Alter	
Score	Freq.	Percent	Freq.	Percent
-1.694785	566	2.65	163	0.81
-1.563741	1,069	5.00	1,596	7.96
-1.523125	209	0.98	210	1.05
-1.488498	867	4.05	755	3.77
-1.295346	1,520	7.10	1,113	5.55
-0.9188861	2,182	10.20	1,696	8.46
-0.7637425	1,072	5.01	1,514	7.55
-0.7290986	381	1.78	355	1.77
-0.6152064	332	1.55	437	2.18
-0.5838171	1,225	5.73	991	4.94
-0.5739842	533	2.49	479	2.39
-0.3990526	120	0.56	459	2.29
-0.2801967	306	1.43	83	0.41
-0.2024503	154	0.72	32	0.16
-0.1149778	1,464	6.84	1,228	6.12
-0.0760427	1,732	8.09	2,138	10.66
0.0658743	544	2.54	702	3.50
0.138291	485	2.27	668	3.33
0.4168512	643	3.01	390	1.95
0.6803553	219	1.02	223	1.11
0.7463204	838	3.92	661	3.30
0.766371	399	1.86	322	1.61
0.8302992	931	4.35	901	4.49
0.8631468	764	3.57	993	4.95
1.028427	96	0.45	143	0.71
1.296386	298	1.39	247	1.23
1.324646	815	3.81	462	2.30
1.369108	810	3.79	533	2.66
1.40581	619	2.89	282	1.41
1.95731	204	0.95	273	1.36
Total	21,397	100.00	20,049	100.00

Annual earnings

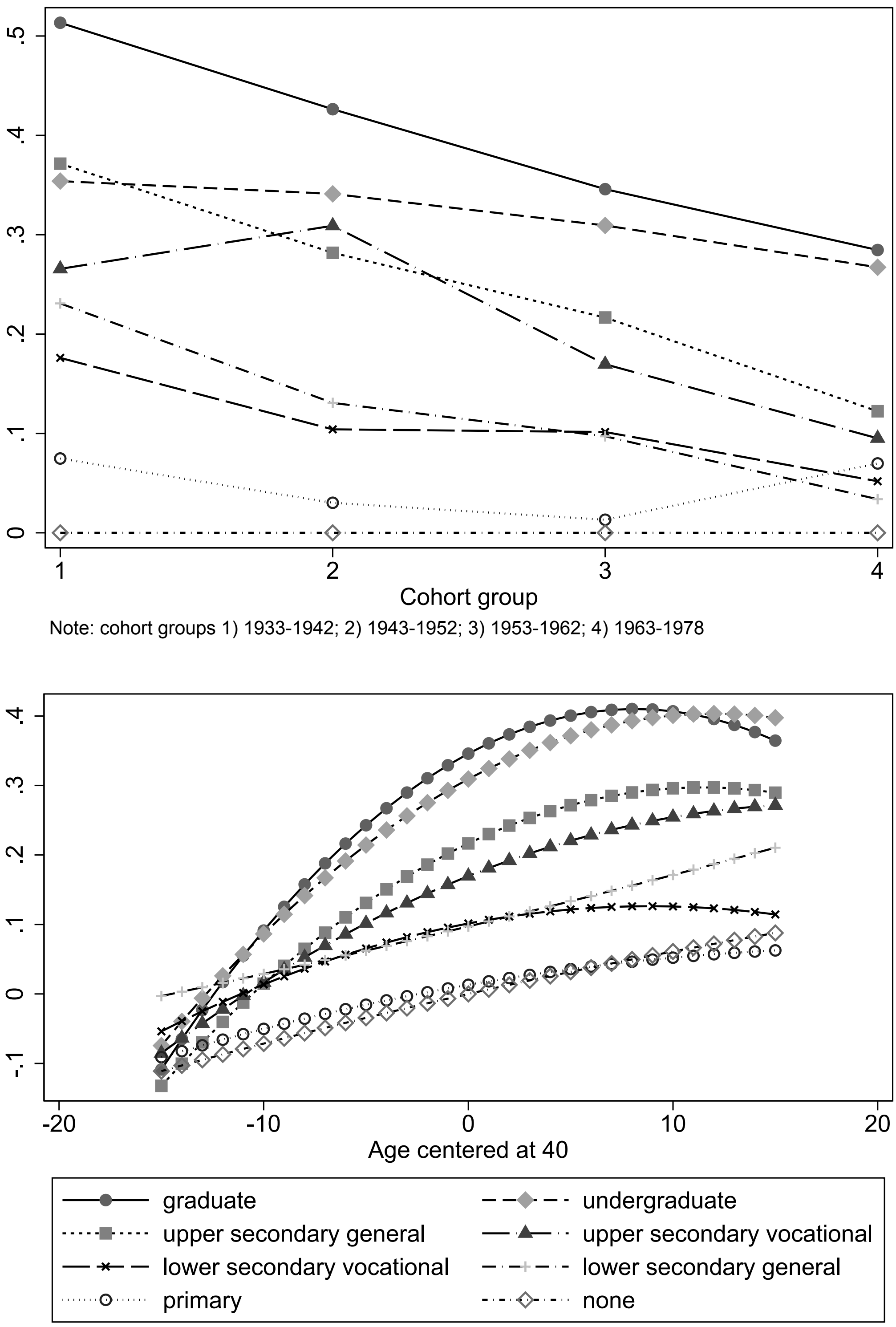


Figure 2.4: Earnings gains by education and cohort with "no degree" as reference, and returns to age by education for the group reference "born 1953-1962", for women

Inference in sibling correlations

Pearson's correlation coefficient is approximatively normally distributed for small absolute values of correlation. However for higher values the distribution is skewed. That is why for inference issues we use the so-called Fisher's z transformation to convert Pearson's ρ to the normally distributed variable z, with the standard error σ_z (and number of observations n):

$$z = \tfrac{1}{2} ln \tfrac{1+\rho}{1-\rho},$$
$$\sigma_z = \tfrac{1}{\sqrt{n-3}}.$$

In order to test whether correlation coefficients from two independent groups 1 and 2 are statistically different:

$$H_0 : \rho_1 = \rho_2$$
$$H_1 : \rho_1 \neq \rho_2,$$

we compute the test statistic U, following the standard normal distribution under the null hypothesis:

$$U = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}.$$

Chapter 3

Instrumenting education in France: Using May 1968 events as a natural experiment?

3.1 Introduction

Economists have long been interested in the causal effect of education, both when investigating its returns on the labor market and the intergenerational transmission of inequality. They have been confronted to an endogeneity issue, as education is correlated with unobservable characteristics also affecting the variable of interest, income or children's outcome. In particular not taking into account the ability-bias yields OLS estimates to be upwardly biased (Blackburn and Neumark, 1993). In order to control for this endogeneity bias, one approach is to exploit exogenous sources of variation in education deriving from natural experiments.

Regarding returns to education, Angrist and Krueger (1991, 1992) as well as Leigh and Ryan (2008) use age at school entry and compulsory schooling laws.

Minimum school leaving age (MSLA) reforms are also used by Acemoglu and Angrist (1999), Meghir and Palme (2005), Oreopoulos (2006) and Aakvik, Salvanes, and Vaage (2010). Alternatively, Ichino and Winter-Ebmer (2004) exploit children's educational loss caused by their father being at war and Gurgand and Maurin (2007) the strong educational expansion after WWII in France. Butcher and Case (1994) use the number and sex composition of the siblings and Duflo (2000) exploits school construction. For the intergenerational transmission of inequality, MSLA reforms are widely used, in particular by Chevalier (2004), Black et al. (2005), Oreopoulos et al. (2006), Holmlund (2006) and Holmlund et al. (2011). Alternatively Carneiro et al. (2013) use changes in school costs and Currie and Moretti (2003) the availability of colleges.

Concerning the use of natural experiments, issues have been highlighted since Angrist and Krueger (1991) received many critics about their instrumental method, in particular from Bound and Jaeger (1996, 2000). Indeed as seen in Bound, Jaeger, and Baker (1995), when using a weak instrument even a small correlation between the instrument and the error term in the original estimation can lead to large inconsistencies in the IV estimates. In any case, Bound, Jaeger, and Baker (1993); Bound et al. (1995) and Staiger and Stock (1997) strongly recommend to always provide the F-statistics associated with the first stage estimations to assess the quality of the instrument. As an additional concern, contrary to random control trials, in natural experiments the treatment and control groups may not be comparable even if random assignment is assumed, as detailed in Sekhon and Titiunik (2012), which invalidates the identification.

In France, Grenet (2013) and Maurin and McNally (2008) use natural experiment methods to instrument education. Grenet (2013) uses the increase from 14 to 16 years old of the MSLA induced by the Berthoin Law to estimate the returns to education. Maurin and McNally (2008) exploit the events of May 1968 to investigate both returns to education and intergenerational mobility. During spring

1968, a wave of student protests escalated into a general strike. It led among others to important modifications in examination modalities, especially for the high school certification, called *baccalauréat* in France. Consequently more high schoolers graduated this year and thus had the opportunity to have access to higher education. The authors use these modifications in examination modalities following the events of May 1968 in France to investigate the returns to education and the intergenerational transmission of education. They argue that "unlike all other papers in the literature, the intervention is a one-off, unexpected, and temporary: it has no consequences for cohorts coming after 1968 events, and the incentive structure of the educative system is unchanged" and conclude that the events thus fit the prerequisites of a convincing natural experiment.

However we have some concerns regarding the nature of the events of May 1968 and the context in which they happened, as well as more specifically the strategy implemented in Maurin and McNally (2008). First of all we suspect a slight impact of the events on education, since only a limited number of students who would not have passed otherwise graduated from high school thanks to the events. Indeed for most of students there was no effect, neither on low-performing students who even in this context did not obtain the *baccalauréat*, nor on high-performing students who would have graduated anyways. Furthermore the modification of the examination modalities were not the only change occurring at this time. A new type of *baccalauréat* was created in 1968 (with a first session in 1969) as an alternative to the existing *baccalauréat général*: the *baccalauréat technologique*. The former consists of general studies, whereas the latter is more job-oriented. This new orientation possibility changed the composition of the population of university students, as high schoolers now had the opportunity to select a shorter, vocational path (see Cappellari and Lucifora (2009) for a change in the higher education structure in Italy).

Regarding the specific method of Maurin and McNally (2008), our first doubt concerns the instrumental variable. They choose the birth cohort of 1949 as the population treated by the events of May 1968, since the median age of *baccalauréat* candidates was 19 at this time (the standard age to take the examination is 18, but half of pupils repeated at least one grade in primary school). However this measure does not accurately target the affected population, since part of individuals born in 1949 took the examination another year and individuals born years prior or subsequent to 1949 took the examination in 1968 (in particular individuals born in 1950, who were candidates at age 18). Besides Maurin and McNally (2008) find IV estimates much higher than the OLS ones. On wages, the ratio is about 1.5 and on children's education (grade repetition) more than 4. About this last finding, the authors argue that their "results are qualitatively similar to Oreopoulos et al. (2006) in that larger effects are estimated when using the IV approach". In the cited study, IV estimates are however only almost twice the size of the OLS ones. All of the above raises weak instrument concerns, which can not be dismissed as the authors do not extensively discuss the first stage estimations, nor the F-statistics.

Our contribution is first to reveal that the variable used in Maurin and McNally (2008) – being born in 1949 – is not a valid instrument for education. Second and more generally, we show that the events of May 1968 – even when represented by the year of *baccalauréat* examination – do not qualify as a cogent natural experiment. We start by replicating the estimations of Maurin and McNally (2008), using the same empirical procedure on Labor Force Survey (LFS) data. We report first stage estimates and F-statistics as well as OLS and IV results, as in Maurin and McNally (2008), for the main strategy and for alternative specifications. The results obtained confirm our weak instrument concerns. To check whether this is due to small sample size, we reproduce the same estimations on Census data. However the instrument fails at the placebo tests. Since we excluded the use of birth year 1949 as an instrument, we move on to the year of *baccalauréat* examination, available in the Education-Training-Employment (FQP) survey. Indeed even though being born in 1949 proves to be an unsuitable instrument, the events of May 1968 could still be used as a natural experiment, and having taken the *baccalauréat* examination in year 1968 is a better reflection of having been affected. Nonetheless, using FQP data, we show that since students had the opportunity to take the examination more than once by repeating grades, the events only increased the rate of success in year 1968 but not the final level of education of the treated population. All in all we conclude that the events of May 1968 do not constitute a relevant natural experiment to instrument education.

The remainder of this paper is organized as follows. Section 3.2 presents the events of May 1968 and the context in which they took place, as well as the changes they brought to the French educational system. Section 3.3 explains the strategy implemented, through the estimation and sampling procedures and the description of the different data sets used (LFS, Census and FQP). Section 3.4 presents the replication results obtained from LFS data and Section 3.5 the comparable results obtained from alternative data sets. Finally, we conclude in Section 3.6.

3.2 May 1968 events: context and aftermath

In order to instrument education, Maurin and McNally (2008) use the events of May 1968 as a natural experiment, to obtain an exogenous variation in education. In France, during spring 1968, a general strike initiated by a wave of student protests paralyzed the entire country. It all started with the protestation of a small group of students reacting to the arrest of fellow students during an anti-Vietnam War demonstration. It then rapidly escalated into a massive student riot centered around student condition considerations. Following the violent repression of the movement by police forces – yielding hundreds of severely injured people – the main labor unions decided to join the protest and called a general strike. About 10 million workers stopped working, factories and universities were closed or occupied, leading to negotiations with the government.

The movement took an end with the signature of agreements taking account of labor force and student demands. The evaluation conditions for high school and university examinations were a focal point of the student negotiations. Indeed this school year had been profoundly disturbed for all students, involved or not in the revolt. Many universities decided to delay and/or revise their examinations, but certainly the most important modification affected the national examination of the *baccalauréat*. This high school certification, taking place every year in June, ends the secondary education and guarantees access to universities, so to speak free of charge in France.

The negotiations following the events of May 1968 led to a half day of only oral examinations, instead of the usual week-long almost all-written examinations. Additionally, high school students were given their results right after the examination, hence examiners could not coordinate with one another to bring grades into line, as the procedure normally requires. In all likelihood these changes affected students only positively, among other things because professors relied more on school reports.[1]

As a result, for the entire student population (men and women) in 1968 the rate of success was almost 20 percentage points higher (59.6% in 1967, 81.3% in 1968, and 66.0% for all *baccalauréats* and 67.6% for the *baccalauréat général* in 1969[2]) and the proportion of *bacheliers* in a generation – as defined by the Ministry of Education – was around 4 percentage points higher (15.4% in 1967, 19.6% in 1968, and 16.1% for all *baccalauréats* and 14.4% for the *baccalauréat général* in 1969).[3] Indeed in Figure 3.1 displaying the proportion of *bacheliers* in a generation, one can see a peak for year 1968. However if the proportion of *baccalauréat général* comes

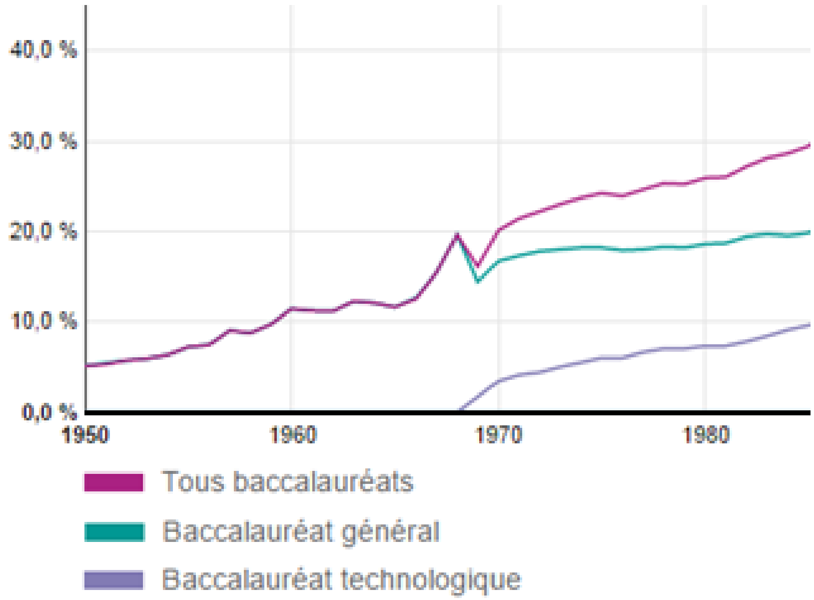
[1]School reports constitute a decision-making tool for the *baccalauréat* jury. Filled by professors during school year, they record a student's knowledge and progress. In particular they can be used during deliberations for students whose examination results are just below the admission threshold.

[2]Technological *baccalauréats* were created in 1968 and the first session took place in 1969, as seen in Figure 3.1. The original *baccalauréat* became the *baccalauréat général* as it consists of general studies. The technological *baccalauréat* (*baccalauréat technologique*) however is more job-oriented.

[3]Source: French Ministry of Education.

back to its value of 1967 in 1970, the proportion of all types of *baccalauréats* is already as high as in 1968 in 1970.

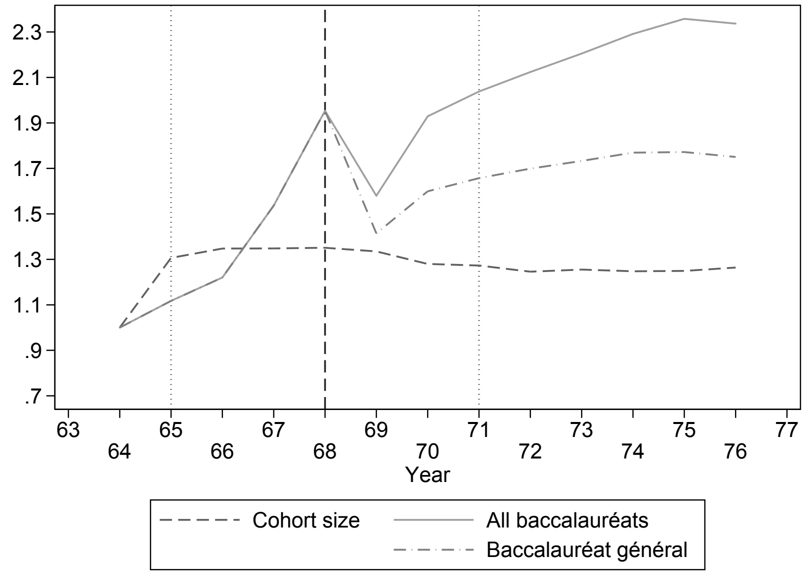
Figure 3.1: Proportion of *bacheliers* in a generation



- Note: *"Tous baccalauréats"* stands for "All types of high school certificates", *"Baccalauréat général"* for "General high school certificate" and *"Baccalauréat technologique"* for "Technological high school certificate".
- Definition of the proportion of *bacheliers* in a generation: proportion of *bacheliers* of a fictive generation of individuals who would have, for each age, the participation and success rates observed the considered year. This number is obtained by calculating for each age the proportion of *baccalauréat* owners in the total population of this age, and by summing these rates by age.
- Source: "07. Le baccalauréat et les bacheliers" in "L'état de l'Enseignement supérieur et de la Recherche en France n°9 - June 2016", Ministry of Education.

Maurin and McNally (2008) display the trends in the number of *bacheliers*, again for men and women, using data from the French Ministry of Education, and in cohort size, using data from the French Statistical Office INSEE (see Figure 3.7 in appendix, extracted from Maurin and McNally (2008)) by year of examination (year t corresponding to birth year $t-19$ as 19 is the median age for the candidates). They observe a clear and unique peak for year 1968, with a rate of *bacheliers* returning just after the events to its preceding value. Overall the events of May 1968, and more specifically the modification of the examination's modalities this year, seem to provide an interesting source of exogenous variation in education. As such Maurin and McNally (2008) conclude that it constitutes a favorable framework to implement natural experiment methods.

Figure 3.2: Trends in the number of individuals passing the *baccalauréat* and *baccalauréat général* and in cohort size



Source: French Ministry of Education (number of *baccalauréats*) and French Statistical Office INSEE (cohort size). As in Maurin and McNally (2008), the size of the cohort for year t corresponds to the number of persons born at $t - 19$ (19 is the median age for the candidates) and the two series are normalized to 1 for 1964. This Figure corresponds to Figure 1 in Maurin and McNally (2008).

To start off the replication of the study of Maurin and McNally (2008), we plot in Figure 3.2 the trends in cohort size and number of *bacheliers* by year of examination, for both men and women, using the same specification and data source (see Table 3.11 in appendix for detailed statistics about the *baccalauréat* examination year by year, from 1945 to 1953). First, we thus consider all kinds of *baccalauréats*, but do not find the same pattern as the authors. The proportion of *bacheliers* in 1970 is already back to the value of 1968, which does not correspond to the findings of Maurin and McNally (2008), for whom the value in 1970 is close to the one of 1967. The solid line in Figure 3.2 follows the same pattern as the proportion for "Tous baccalauréats" in Figure 3.1, whereas the proportions depicted in Figure 3.7 seem a lot closer to the pattern for "Baccalauréat général". Therefore we also depict the trends in the number of *baccalauréat général*, the dashed line in Figure 3.2. The shape of this dashed line indeed looks like the one for "Baccalauréat général" in Figure 3.1, as the one of Maurin and McNally (2008).

It seems that Maurin and McNally (2008) actually show the trends for the *baccalauréat général* and not for all kinds of *baccalauréat*, which is however not specified in their paper. As the peak for the *baccalauréat général* corresponds more to a temporary event (not affecting the following years), one can wonder whether it constitutes a better educational variable of interest than all kinds of *baccalauréat*. However from 1969 – year of the first session of the *baccalauréat technologique* – people who otherwise would have taken a (*général*) *baccalauréat*, as well as people who otherwise would not have taken the *baccalauréat*, have the opportunity to choose a *baccalauréat technologique* instead. Thus it is not clear to which educational variable the situation prior to 1969 should be compared: *baccalauréat général* or all kinds.

Additionally as opposed to what is argued in Maurin and McNally (2008), the events of May 1968 are not temporary (at least concerning all kinds of *baccalauréats*), they had consequences for cohorts coming after 1968 and the incentive structure of the educative system changed, in particular due to the creation of the *baccalauréat technologique* and the Faure law.[4] These considerations question the validity of the events of May 1968 as a favorable framework to implement natural experiment methods. To our understanding and based on the descriptive statistics presented, Maurin and McNally (2008) seem to include all types of *baccalauréats* in their estimations (and not only the *baccalauréat général*). We will both replicate their analysis and investigate the differences observed using either all kinds of *baccalauréats* or only the *baccalauréat général*.

[4]The Faure law was passed in the aftermath of the events, on November 1968, when Edgar Faure was Minister of Education. It shifted the role played by higher education in France, placed more emphasis on formation and training of universities' students, gave greater autonomy to the universities and yielded a democratic management of the university.

3.3 Method and data

3.3.1 Estimation strategy

To evaluate whether the events of May 1968 had an impact on education and thus qualify as a natural experiment to instrument it, we consider the following first stage equation:

$$Y = \alpha_1 May68 + \alpha_2 X + \epsilon,$$

where Y represents education level, and X is a set of control variables. We use different educational outcomes: having a *baccalauréat*, having a university diploma or degree and the number of years of higher education. Indeed not only the success at the end of high school but also the following higher education should be affected by the events of May 1968, maybe not to the same extent. Since French universities do not select first year students and are nearly free of charge, graduating from high school guarantees the opportunity to access higher education.

$May68$ is a dichotomous variable equal to 1 if the individual is affected by the events, 0 if not. An individual is considered affected by the events of May 1968 if he/she took the *baccalauréat* examination in 1968. However, the year a student takes this examination is not available in all datasets, but only in data from FQP surveys (among the datasets we use). Thus with FQP data, we can use the year a student takes the *baccalauréat* examination as an instrumental variable for education.

Alternatively, using data from each dataset (LFS, Census and FQP for comparative reasons), we use birthyear 1949 as the instrument, as in Maurin and McNally (2008). The median age at which the *baccalauréat* examination was taken in France is 19 years old.[5] This is why we consider as affected by the events of May 1968 people born in 1949. Moreover, we do not particularly expect high performing students to be affected by the events, as they would have passed without any relaxation of

[5]French students start school at age 6 and complete 12 years of education by the end of high school so that the standard age to take the *baccalauréat* examination is 18. However at that time half of the student population repeated at least one grade during primary education.

the examination conditions. We anticipate a greater impact of the events on less performing students, typically those who repeated a grade.

3.3.2 Samples and descriptive statistics

Replication of Maurin and McNally (2008)

To perform the replication exercise, the first dataset used here comes from the Labor Force Survey (LFS), as in Maurin and McNally (2008). The LFS is a nationally representative sample of individuals aged 15 and above. Maurin and McNally (2008) use only the waves 1990, 1993, 1996 and 1999 of the LFS since the sample rotates every three years (the same individuals are thus interviewed three years in a row) and they want to observe each individual only once. They justify using data starting from 1990 by the need of information on wages (to estimate the returns to education), only available from this time onwards. They focus on male workers born between 1946 to 1952. We apply the same sampling strategy and report comparable descriptive statistics in Table 3.1, as well as for all men.

Different education outcomes are considered. Regarding education dummies, "Less than *baccalauréat*" corresponds to individuals who do not hold a high school degree, "*Baccalauréat* only" corresponds to individuals who only hold a high school degree but no higher education, a "University diploma" corresponds to a higher education diploma obtained two years after the *baccalauréat* and a "University degree" corresponds to any higher education degree obtained minimum three years after the *baccalauréat*. The number of years of higher education is not directly available in the LFS, so we construct this variable based on the highest degree obtained by the individual, following Maurin and McNally (2008): "Years of higher education" equals 0 for "Less than *baccalauréat*", 3 for "*Baccalauréat* only", 5 for "University diploma" and 7 for "University degree".

Overall our sample of 26,293 male workers seems very similar to the one of 26,371 male workers in Maurin and McNally (2008). The repartition in the birth cohort

Table 3.1: Descriptive statistics in Maurin and McNally (2008) and replicated using LFS

	Maurin & McNally		**Baguet & Lecavelier**			
	Male wage earners		Male wage earners		All men	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Cohort dummy						
1946	0.128	(0.33)	0.128	(0.33)	0.133	(0.34)
1947	0.140	(0.35)	0.140	(0.35)	0.144	(0.35)
1948	0.145	(0.35)	0.146	(0.35)	0.148	(0.35)
1949	0.148	(0.35)	0.148	(0.35)	0.146	(0.35)
1950	0.145	(0.35)	0.145	(0.35)	0.145	(0.35)
1951	0.145	(0.35)	0.145	(0.35)	0.142	(0.35)
1952	0.148	(0.35)	0.148	(0.35)	0.143	(0.29)
Education dummy						
Less than *Baccalauréat*	0.718	(0.45)	0.722	(0.45)	0.728	(0.44)
Baccalauréat only	0.096	(0.29)	0.097	(0.30)	0.094	(0.29)
University diploma	0.074	(0.26)	0.073	(0.26)	0.068	(0.25)
University degree	0.111	(0.31)	0.108	(0.31)	0.111	(0.31)
Years of higher education	1.440	(2.47)	1.408	(2.45)	1.398	(2.45)
Wage (log)	9.170	(0.43)	9.176	(0.50)		
Observations	26,371		26,293		36,629	

Source: LFS 1990, 1993, 1996 and 1999.
Sample: Male wage earners/men born between 1946 and 1952.
Specification: "Years of higher education" equals 0 for "Less than *baccalauréat*", 3 for "*Baccalauréat* only", 5 for "University diploma" and 7 for "University degree".
Replication: This table corresponds to Table 1 in Maurin and McNally (2008).

groups is equivalent in both samples. Concerning the education level, the repartition is also the same, at 0.4 percentage points at most. We find a slightly lower average and standard deviation of years of higher education (1.440 (2.47) in Maurin and McNally (2008) versus 1.408 (2.45) here). The average and standard deviation of log wage are slightly higher in our sample (9.176 (0.50) versus 9.170 (0.43)). As for the sample of 36,629 men (not only wage earners), individuals with less than a *baccalauréat* or at least a university degree are more represented. Indeed among non-wage earners we now include in the sample some high education professions (for instance private practice doctors, lawyers, architects, ...), but also many lower education occupations (for instance craftspeople, storekeepers, farmers, ...). As a result the average number of years of higher education is slightly lower.

Due to the creation of the *baccalauréat technologique* in 1969, we expect the effect of the events of May 1968 to be stronger on the *baccalauréat général*. Thus we also construct two alternative educational dummy variables: equal to 1 if the individual has at least a *baccalauréat* – either any kind or a *baccalauréat général* – 0 otherwise. The descriptive statistics relative to these variables are presented in Table 3.12 in appendix.

Alternative datasets: Census and FQP

To further investigate the validity of the events of May 1968 as an instrument for education, we use Census data to observe the French population as a whole, as we expect only a small proportion of the population to have been affected. The second dataset used in this paper comes from the French Census of 1999, which is the last wave of entire population survey in France. After that, the French Census becomes an annual survey from 2004 on, with approximately 8% of housing surveyed each year. This database contains both birth year and education level, which allows us to perform our first stage regression on a large sample.

Census data do not contain any information neither on wage nor on the year students have taken their examinations. Thus we select men born between 1946 and 1952 and use birth year 1949 as our instrument. The educational outcomes are the same as in Maurin and McNally (2008) and in our replication using LFS, in Section 3.3.2.

We also use data from the FQP survey, which targets 18 to 65 years old individuals. As a detailed educational calendar is provided since the survey of 1993, we use the waves 1993 and 2003. We do not exploit the wave 2014 as surveyed individuals are too young to be part of our sample. The year students take their examinations being available, we are able to assess the impact of taking the *baccalauréat* examination in 1968 on the probability of success for men born between years 1946 and 1952. Additionally, for comparison purposes we again evaluate the effect of birth year 1949 on education, investigating the same educational outcomes as with LFS and Census data.

For both Census and FQP datasets, descriptive statistics for the two alternative educational dummy variables – at least a *baccalauréat* and at least a *baccalauréat général* – are reported in Table 3.12 in appendix.

Table 3.2: Descriptive statistics using Census and FQP

	Census		**FQP**	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Cohort dummy				
1946	0.137	(0.344)	0.145	(0.352)
1947	0.143	(0.351)	0.147	(0.354)
1948	0.145	(0.352)	0.158	(0.365)
1949	0.146	(0.353)	0.139	(0.346)
1950	0.146	(0.353)	0.150	(0.357)
1951	0.141	(0.348)	0.128	(0.334)
1952	0.142	(0.349)	0.134	(0.340)
Education dummy				
Less than *Baccalauréat*	0.699	(0.459)	0.751	(0.432)
Baccalauréat only	0.113	(0.317)	0.105	(0.306)
University diploma	0.071	(0.256)	0.050	(0.217)
University degree	0.117	(0.322)	0.094	(0.217)
Years of higher education	1.514	(2.499)	1.223	(2.304)
Observations	2,488,383		4,828	

Source: Census 1999 and FQP 1993 and 2003.

Sample: Men born between 1946 and 1952.

Specification: "Years of higher education" equals 0 for "Less than *baccalauréat*", 3 for "*Baccalauréat* only", 5 for "University diploma" and 7 for "University degree".

As reported in Table 3.2, the sample constructed with Census data contains 2,488,383 men. Among these men, 69.9% have less than a *baccalauréat*, 11.3% have only a *baccalauréat*, 7.1% have a university diploma and 11.7% a university degree. The corresponding rates for the FQP sample of 4,828 men are 75.1%, 10.5%, 5.0% and 9.4% respectively. The rate of men without a *baccalauréat* is lower in the Census sample than in LFS and FQP samples, whereas the rates for all other education dummies are higher. When comparing the FQP sample to the LFS sample, the rates for a *baccalauréat* or less are higher and the rates for more than a *baccalauréat* are lower. Additionally, in the FQP sample of 1,130 men who ever took the *baccalauréat* examination, 87.7% succeeded in obtaining it (either at first try or by repeating the last year of high school).

3.4 Replication of Maurin and McNally (2008) using the LFS

3.4.1 Impact of birth year 1949 on education and wage

In Section 3.2, we presented in Figure 3.2 the proportion of *bacheliers* by year of examination using data from the French Ministry of Education. Following the idea that being born in 1949 is a good representation of taking the *baccalauréat* examination in 1968 (and thus being affected by the events of May 1968), we display in Figure 3.3 the proportion of men who have at least a *baccalauréat* by birth year, from 1939 to 1959, using LFS data. Again we distinguish all men from wage earners. There is an increasing trend and a small – but not unique – peak for birth year 1949. Besides the peak is lower when considering all men and not only wage earners.

Figure 3.3: Proportion of individuals passing the *baccalauréat* among all men and male wage earners born from 1939 to 1959 using LFS

Source: LFS 1990, 1993, 1996 and 1999.

Maurin and McNally (2008) start by assuming that education varies in a non-linear way across birth cohorts. They regress educational outcomes and log wages on cohort dummies, for a sample of male workers born between 1946 and 1952. We replicate the same estimations and both sets of results are reported in Table 3.3.

Table 3.3: Impact of birth year on education and labor market outcomes in Maurin and McNally (2008) and replicated using LFS

	First stage				Reduced form
	Baccalauréat only	University diploma at least	University degree at least	Years of higher education	Log wage
	(1)	(2)	(3)	(4)	(5)
Maurin & McNally					
1947	-0.009 (0.006)	0.014 (0.008)	0.008 (0.006)	0.060 (0.050)	0.006 (0.010)
1948	0.007 (0.006)	0.015 (0.008)	0.012 (0.006)	0.080 (0.050)	0.031 (0.010)
1949	-0.001 (0.006)	0.027 (0.008)	0.009 (0.006)	0.150 (0.050)	0.021 (0.010)
1950	0.001 (0.006)	0.008 (0.008)	-0.002 (0.006)	0.030 (0.050)	0.005 (0.010)
1951	-0.005 (0.006)	0.002 (0.008)	-0.001 (0.006)	0.010 (0.050)	0.003 (0.010)
Trend	-0.000 (0.001)	0.001 (0.008)	-0.001 (0.001)	0.005 (0.010)	0.010 (0.002)
Age	-0.000 (0.001)	0.001 (0.008)	-0.000 (0.001)	0.004 (0.005)	0.023 (0.001)
Observations	26,370	26,370	26,370	26,370	26,370
Baguet & Lecavelier					
1947	-0.009 (0.007)	0.014 (0.008)	0.009 (0.007)	0.060 (0.054)	0.003 (0.011)
1948	-0.008 (0.006)	0.014 (0.008)	0.011 (0.006)	0.007 (0.050)	0.035 (0.010)
1949	-0.001 (0.006)	0.027 (0.008)	0.009 (0.006)	0.150 (0.049)	0.022 (0.010)
1950	-0.001 (0.006)	0.008 (0.008)	-0.003 (0.006)	0.029 (0.050)	0.006 (0.010)
1951	0.001 (0.006)	-0.001 (0.008)	-0.003 (0.007)	-0.007 (0.051)	0.004 (0.010)
Trend	-0.000 (0.001)	0.003 (0.002)	0.001 (0.001)	0.014 (0.011)	0.010 (0.002)
Age	-0.000 (0.001)	0.002 (0.001)	0.001 (0.001)	0.011 (0.005)	0.023 (0.001)
R-squared	0.000	0.001	0.001	0.001	0.026
Observations	26,293	26,293	26,293	26,293	26,293

Source: LFS 1990, 1993, 1996 and 1999.
Sample: Male wage earners born between 1946 and 1952.
Specification: Coefficients for birth year dummies 1947 to 1951 are relative to the comparison birth years of 1946 and 1952.
Replication: This table corresponds to Table 4 in Maurin and McNally (2008).

We are able to accurately reproduce the results of Maurin and McNally (2008) for each first stage educational outcome – *baccalauréat* only, at least a university diploma or degree and the number of years of higher education – as well as for log wage, for the reduced form. When our estimates slightly deviate from the ones of Maurin and McNally (2008), they are not significant (neither for our estimations nor for theirs).

Maurin and McNally (2008) then choose to construct an instrumental variable – further referred to as $May68$ – equal to 1 when the individual is born in 1949 and 0 when the individual is born either in 1946 or 1952. They select the symmetrical birth cohorts of 1946 and 1952, arguing that they are less likely affected by the events of May 1968 than the years in between, closer to 1949. As endogenous educational variable, they use the number of years of higher education.

The authors do not discuss the quality of their instrument and do not explicitly report the F-statistics. These F-statistics still can be obtained as they equal the t-statistics squared, in the case of a unique instrument, and the t-statistics can be computed from the estimates and standard errors, however approximated. We computed this effect of $May68$ on having a *baccalauréat* only, having any university diploma or degree and the number of years of higher education, both for wage earners and all men, as seen in Table 3.4.

Table 3.4: First stage estimations: Effect of $May68$ on education outcomes for male wage earners and all men using LFS

	Male wage earners			**All men**		
	Baccalauréat only	University diploma or degree	Years of higher education	*Baccalauréat* only	University diploma or degree	Years of higher education
	(1)	(2)	(3)	(4)	(5)	(6)
$May68$	-0.001	0.027***	0.150***	-0.003	0.022***	0.118***
	(0.006)	(0.008)	(0.049)	(0.005)	(0.006)	(0.042)
Birth year	0.001	0.004**	0.027**	0.001	0.003*	0.020**
	(0.001)	(0.002)	(0.012)	(0.001)	(0.002)	(0.010)
F-stat	0.01	12.34	9.52	0.33	11.01	8.09
R-squared	0.000	0.002	0.002	0.000	0.001	0.001
Observations	11,145	11,145	11,145	15,433	15,433	15,433

Source: LFS 1990, 1993, 1996 and 1999.

Sample: Male wage earners/men born in 1946, 1949 and 1952.

Specification: $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or 1952.

Significance levels : $*$: 10% $**$: 5% $***$: 1%

The results of columns (1) and (4), on *baccalauréat* only, are not statistically significant. Concerning column (2) (resp. (5)), being born in 1949 (versus 1946 or 1952) increases the likelihood of having a university diploma or degree by about 3 (resp. 2) percentage points, which corresponds to a probability of around 21% (resp. 20%), instead of 18%. However the F-statistic is only at 12.34 for wage earners and at 11.01 for all men, above but not far from the admitted threshold for weak instrument (F-statistic at 10).

Columns (3) and (6) respectively show individuals born in 1949 present 0.15 and 0.12 additional years of higher education, compared to individuals born in 1946 or 1952. So the effect of $May68$ on the number of years of higher education is significant but small. However the F-statistic are only 9.52 and 8.09, thus stand below the weak instrument threshold. These results raise the question of the validity of being born in year 1949 as an instrument.

Thus the impact of the instrument on educational outcomes is limited for men. In Figure 3.1, we observed a peak in the rate of *bacheliers* in year 1968 for the entire student population (men and women). We compute the same estimation on the female sample, as reported in Table 3.5. Surprisingly, being born in 1949 versus 1946 or 1952 does not have a significant impact on any educational outcome. This strengthens our doubts regarding the instrument.

Table 3.5: First stage estimations: Effect of $May68$ on education outcomes for women using LFS

	Female wage earners			**All women**		
	Baccalauréat only	University diploma or degree	Years of higher education	*Baccalauréat* only	University diploma or degree	Years of higher education
	(1)	(2)	(3)	(4)	(5)	(6)
$May68$	-0.007	-0.002	-0.038	-0.007	-0.000	-0.027
	(0.007)	(0.009)	(0.053)	(0.005)	(0.006)	(0.038)
Birth year	0.004***	0.005**	0.035***	0.003**	0.007***	0.044***
	(0.002)	(0.002)	(0.013)	(0.000)	(0.001)	(0.009)
F-stat	0.99	0.05	0.53	1.79	0.00	0.48
R-squared	0.001	0.001	0.001	0.000	0.001	0.002
Observations	9,624	9,624	9,624	15,954	15,954	15,954

Source: LFS 1990, 1993, 1996 and 1999.

Sample: Female wage earners/women born in 1946, 1949 and 1952.

Specification: $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or 1952.

Significance levels : $*$: 10% $**$: 5% $***$: 1%

3.4.2 OLS and IV estimations

We pursue the replication exercise by computing the same OLS and IV estimations as in Maurin and McNally (2008): the effect of the number of years of higher education on wages, instrumented by $May68$. Both sets of results are presented in Table 3.6 and are similar.

The IV estimates are 1.5 times larger than the OLS ones. An explanation suggested by Maurin and McNally (2008) is that it might be due to measurement error

Table 3.6: Evaluation of the returns to education in Maurin and McNally (2008) and replicated using LFS

	Maurin & McNally		**Baguet & Lecavelier**	
	Log wage		Log wage	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Years of higher education	0.094	0.140	0.096	0.145
	(0.002)	(0.060)	(0.002)	(0.060)
Birth year	0.010	0.010	0.102	0.009
	(0.002)	(0.002)	(0.002)	(0.003)
Age	0.023	0.023	0.023	0.022
	(0.001)	(0.002)	(0.001)	(0.002)
R-squared	0.25		0.25	
Observations	11,171	11,171	11,145	11,145

Source: LFS 1990, 1993, 1996 and 1999.

Sample: Male wage earners born in 1946, 1949 and 1952.

Specification: $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or 1952.

Replication: This table corresponds to Table 5 in Maurin and McNally (2008).

on the educational variable, which would downwardly bias the OLS estimate. Nevertheless, the literature documenting the extent of measurement error on self-reported educational variables indicates a downward bias of only 10 to 15%[6] (Ashenfelter and Krueger, 1994; Angrist and Krueger, 1999; Card, 2001), which explains – if anything – a small part of the difference between OLS and IV estimates.

They also argue that it is common in the literature on the wage returns to education, referring to Card (2001). However on the one hand Card (1999) reports that "estimated returns to schooling are 20-40% above the corresponding OLS estimates", reviewing studies using institutional changes in the education system as instrument, noticeably below the 50% found here. On the other hand Card (2001) explains that IV estimates are indeed as large or larger than OLS ones, but for instruments affecting the bottom part of the education distribution,[7] which is not

[6] The reliability ratio of the schooling measure, i.e. the fraction of the variance in the self-reported measures of schooling due to true variation in schooling is estimated around 85 to 90%.

[7] Card (2001) suggests that "marginal returns to education among the low-education subgroups typically affected by supply-side innovations tend to be relatively high, reflecting their high marginal costs of schooling, rather than low ability that limits their return to education".

the case with the events of May 1968. The instrumentation is here used to correct an omitted variable bias, which leads to expect smaller IV than OLS estimates. Furthermore based on the first stage estimations, we suspected that the instrument might be weak, which could also explain the results found here.

In order to investigate whether estimations based on alternative educational outcomes would less suffer from weak instrument bias, we additionally estimate the impact of the instrument on having at least – either any or a *général* – *baccalauréat*. As mentioned, we expect being born in 1949 to affect more the probability of having at least a *baccalauréat général*. For the first stage, the results and F-statistics are reported in Table 3.13 in appendix and are similar to the ones corresponding to having at least a university diploma or degree (0.029 and 0.027 for the estimates, 12.34 and 12.02 for the F-statistics). Concerning the second stage, reported in Table 3.14 in appendix, IV estimates are again about 1.5 times larger than the OLS ones.

3.4.3 Alternative specifications and placebo tests

In order to ensure the validity of their estimation, Maurin and McNally (2008) investigate various alternative specifications, which results are provided in Table 3.7. They slightly change the composition of their control group, keeping individuals born in 1949 as the treated group. They only report the IV estimates, and not the first stage coefficients and the F-statistics. We replicate the specifications and provide both first stage and IV results.

Maurin and McNally (2008) find fairly stable results and argue that it confirms the validity of birth year 1949 as an instrument for education. Concerning sample sizes as well as IV estimates, our results are relatively close. However the F-statistics of our first stage estimations range from 3.44 to 7.50, far from the threshold of 10, indicating a weak instrument. Thus when slightly modifying the control groups, the F-statistic substantially decreases, which discredits the robustness of the instrument.

As an additional verification to check the validity of the natural experiment, we

Table 3.7: Instrumental variable effect of years of education: Alternative specifications in Maurin and McNally (2008) and replicated using LFS

Control Groups	1947 and 1951 (1)	1945 and 1953 (2)	1944-47 and 1950-53 (3)	1944-47 (4)	1950-53 (5)
Maurin & McNally					
IV estimates			**Log wage**		
Years of higher education	0.14 (0.06)	0.16 (0.06)	0.13 (0.06)	0.16 (0.11)	0.18 (0.08)
Observations	11,427	10,292	31,520	16,145	19,262
Baguet & Lecavelier					
IV estimates			**Log wage**		
Years of higher education	0.15 (0.05)	0.27 (0.13)	0.17 (0.07)	0.10 (0.09)	0.12 (0.07)
Observations	11,432	10,294	31,530	16,155	19,272
First stage estimates			**Years of higher education**		
$May68$	0.12 (0.05)	0.09 (0.05)	0.12 (0.04)	0.16 (0.08)	0.17 (0.07)
F-stat	6.52	3.44	7.50	3.79	7.23

Source: LFS 1990, 1993, 1996 and 1999.

Sample: Male wage earners born between 1944 and 1953 depending on specifications.

Specification: $May68$ equals 1 if the individual is born in 1949, 0 for the control groups. Control for age and cohort trend are included in all regressions.

Replication: This table corresponds to Table B1 in Maurin and McNally (2008).

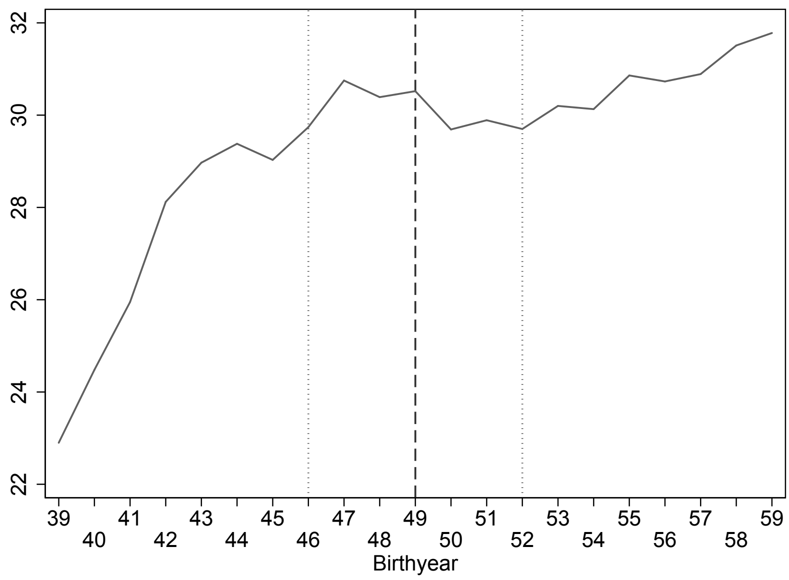
also implement placebo tests for men born between 1939 and 1959, for the number of years of higher education. We use the same "three years before/three years after" rule as in the main specification (1949 versus 1946 and 1952). For example if 1957 is considered as the treated year, then the control group includes 1954 and 1960 cohorts. The results are displayed in appendix in Figure 3.8. Even if the main specification is indeed the one providing the F-statistic closest to the threshold for weak instrument, other placebo specifications yield effects of similar magnitude.

3.5 Alternative data sets

3.5.1 Census Data

In the previous section, we revealed that birth year 1949 is a weak instrument for education using LFS, in the attempt to use the events of May 1968 as a natural experiment. Since it could be related to small sample size issues, we compute the same first stage estimations using Census data to verify whether a larger alternative dataset would yield better results. Indeed, as mentioned, we only expect a small proportion of the individuals to be affected by the instrument: only the birth cohort 1949 and among them only high schoolers who obtained the *baccalauréat* thanks to the events. First we display in Figure 3.4 the proportions of men who have at least a *baccalauréat* by birth year, from 1939 to 1959. There is an increasing trend, however even if the rate of birth year 1949 is higher than the mean of the rates of birth years 1946 and 1952, we do not observe any clear peak for birth year 1949.

Figure 3.4: Proportion of individuals passing the *baccalauréat* among men born from 1939 to 1959 using Census



Source: Census 1999.

To compute the first stage estimate of the effect of birth year 1949 on education, we select as comparison group the symmetrical birth cohorts of 1946 and 1952: the first stage specification used by Maurin and McNally (2008) to compute their IV estimates. Thus, we use the dummy variable $May68$ equal to 1 for individuals born in 1949 and to 0 for those born either in 1946 or 1952. Table 3.8 provides the first stage coefficients for the three educational outcomes: having only a *baccalauréat*, any university diploma or degree and the number of years of higher education. Maurin and McNally (2008) also present these results using Census data[8] in the appendix of their paper and we also report them in Table 3.8. See Table 3.13 in appendix for our results on having at least the *baccalauréat* and at least the *baccalauréat général*.

[8]Maurin and McNally (2008) use a 25% random sample of the 1982 Census, whereas we use

Table 3.8: First stage estimations: Effect of $May68$ on education outcomes in Maurin and McNally (2008) and replicated using Census

	Maurin & McNally			**Baguet & Lecavelier**		
	Baccalauréat only	University diploma or degree	Years of higher education	*Baccalauréat* only	University diploma or degree	Years of higher education
	(1)	(2)	(3)	(4)	(5)	(6)
$May68$	-0.001	0.007***	0.042***	-0.002***	0.009***	0.054***
	(0.001)	(0.001)	(0.008)	(0.001)	(0.001)	(0.005)
Birth year	0.001***	0.000	-0.001	0.000	-0.000	-0.004
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)
F-stat	-	-	-	8.20	141.00	110.26
R-squared	-	-	-	0.000	0.000	0.000
Observations	328,916	328,916	328,916	1,056,384	1,056,384	1,056,384

Source: 25% random sample of Census 1982 for (1), (2) and (3), Census 1999 for (4), (5) and (6).
Samples: Men born in 1946, 1949 and 1952.
Specification: $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or1952.
Replication: This table corresponds to Table C1 in Maurin and McNally (2008).
Significance levels : $*$: 10% $**$: 5% $***$: 1%

We find similar even if slightly higher results as in Maurin and McNally (2008) for all educational outcomes. The effect of the instrument is negative on the likelihood to have only a *baccalauréat*, not significant in Maurin and McNally (2008) contrary to our result, but negligible in both cases. Being born in 1949 – i.e. being affected by the events of May 1968 – increases the probability of having a university diploma or degree by 0.7 to 0.9 percentage points. For instance, for men born in 1949 this probability is 19.7%, versus 18.8% otherwise. Individuals affected by the instrument have 0.04 to 0.05 additional years of higher education.

Thus we still observe a highly significant, but really small effect of $May68$ on higher education. Compared to the coefficients obtained using LFS, the ones from Census data are 3 times smaller for university diploma or degree and number of years of higher education when considering wage earners and 2 times when considering all men. As discussed previously, using Census data to compute these first stage estimates allows us to work with very large samples (almost 100 times larger than the LFS ones) which yields much higher F-statistics, largely above the threshold for weak instruments for these two outcomes.

the full 1999 Census.

To explain the important gap between the results obtained with the two data sets, Maurin and McNally (2008) argue that the Census is "a lot less reliable than the LFS for measuring individual characteristics (notably education and date of birth)". Another source of divergence could have been that the Census does not contain any information on wages. Thus the estimates are computed on men and not on male workers only. Nonetheless the same estimations on all men with LFS (see Table 3.4) yield results lower than on wage earners but still higher than the ones obtained from Census data.

As mentioned, the Census dataset does not contain wage information. We considered a TS2SLS strategy, by computing the first stage on Census data and the instrumentation on LFS data. However the Census first stage estimates are even lower than the LFS ones and thus yielded even higher IV estimates than found in Maurin and McNally (2008).

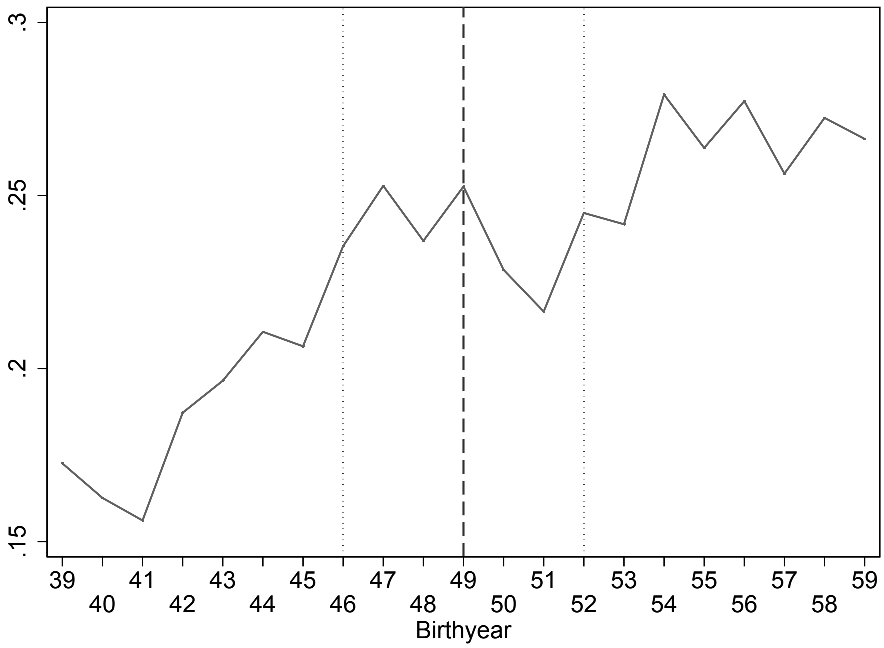
We compute the same placebo tests as in Section 3.4.3 on Census data for the number of years of higher education. The results are displayed in appendix in Figure 3.9. One can see that the results are rather erratic, as we observe 4 specifications with a higher effect than our main specification. Moreover, as opposed to the placebo tests on LFS data, 12 out of the 21 specifications present a F-statistic above 10, which constitutes the conventional limit for weak instrument. Essentially, $May68$ does not stand out as being the only source of variation in education over this time period.

3.5.2 Education-Training-Employment Survey (FQP)

Results provided in Sections 3.4 and 3.5.1 invalidate the use of being born in 1949 as an instrument for education. Nevertheless, the actual instrument is being affected by the events of May 1968, which is more accurately represented by the fact of having taken the *baccalauréat* examination in year 1968, than by being born in 1949. The year individuals have taken their examination is available neither in LFS nor in Census datasets. However, the information is provided in FQP surveys, as a detailed educational calendar is completed for each individual.

For comparison purposes with the other data sources, we display the same representation of the proportion of *bacheliers* by birth cohorts in Figure 3.5. We observe the same increasing trend and a small peak for the birth year 1949, which however does not stand out, as the curve presents other peaks over the time period.

Figure 3.5: Proportion of individuals passing the *baccalauréat* among men born from 1939 to 1959 using FQP



Source: FQP 1993 and 2003.

Again to be able to compare results from FQP surveys with the ones obtained from LFS and Census data, we compute and report in Table 3.9 the same first stage estimations using being born in 1949 as an instrument. We also want to estimate the effect of the alternative instrument, having taken the *baccalauréat* examination in 1968. We then consider samples of men born between 1946 and 1952. Indeed we cannot exclude birth years close to 1949 as we did so far since a substantial proportion of individuals born these years took the examination in 1968 and should be included in the sample. Thus the instrumental variable based on the year an individual took the *baccalauréat* examination is equal to 1 when the individual took this examination in year 1968, 0 if he/she took this examination any other year.

Table 3.9: First stage estimations: Effect of $May68$ on education outcomes using FQP, with either birth year or year of examination as an instrumental variable

	Birth year 1949			**Year of examination 1968**		
	Baccalauréat only	University diploma or degree	Years of higher education	*Baccalauréat* only	University diploma or degree	Years of higher education
	(1)	(2)	(3)	(4)	(5)	(6)
$May68$	-0.023	0.026	0.114	0.003	0.010	0.006
	(0.015)	(0.017)	(0.110)	(0.036)	(0.038)	(0.179)
Birth year	0.001	-0.002	-0.006	0.000	-0.015**	-0.092***
	(0.003)	(0.003)	(0.021)	(0.007)	(0.007)	(0.034)
F-stat	2.40	2.34	1.08	0.01	0.06	0.00
R-squared	0.001	0.001	0.001	0.000	0.004	0.006
Observations	2,015	2,015	2,015	1,130	1,130	1,130

Source: FQP 1993 and 2003.
Sample: For (1), (2) and (3), men born in 1946, 1949 and 1952; for (4), (5) and (6), men born between 1946 and 1952 who took the *baccalauréat* examination.
Specification: For (1), (2) and (3), $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or 1952; for (4), (5) and (6), $May68$ equals 1 if the individual took the *baccalauréat* examination in 1968.
Significance levels : $*$: 10% $**$: 5% $***$: 1%

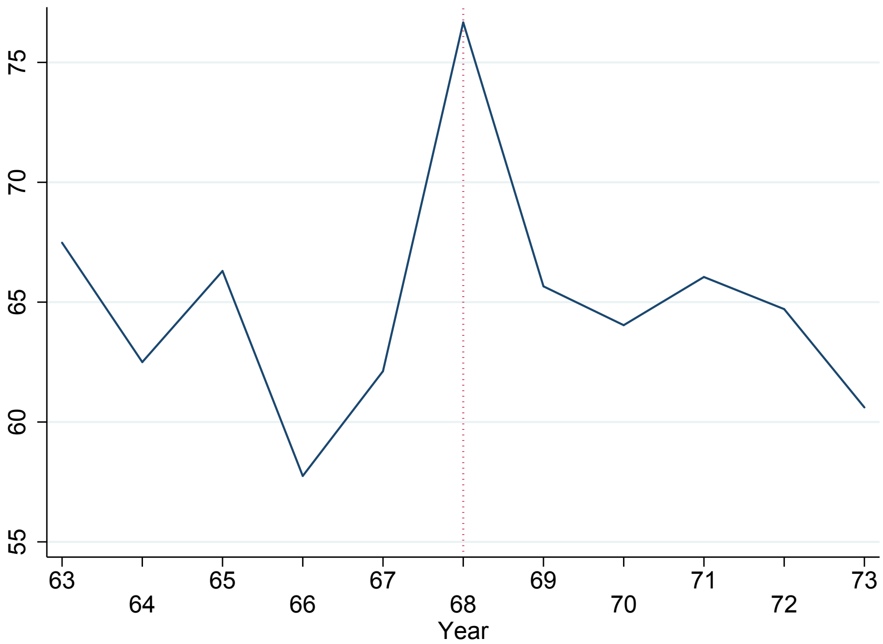
One downside of using the FQP surveys is the sample size. Our FQP sample using birth year 1949 as an instrument consisting of 2,015 men is 5 times smaller than the LFS sample and 500 times smaller than the Census sample. The effect on having only a *baccalauréat* is still negative. The estimates for having a university diploma or degree and number of years of higher education are similar in magnitude to the ones obtained with the LFS data. Additionally, we compute the same placebo tests as for LFS and Census on FQP data for the number of years of higher education. The results are displayed in appendix in Figure 3.10.

The sample used to assess the impact of the year of examination consists of 1,130 men born between 1946 and 1952 who took the *baccalauréat* examination at some point. The estimates are negligible, whichever educational outcome considered. Overall none of our estimates obtained from FQP data are significant and all F-statistics are close to 0 (2.40 at most). Thus neither using birth year 1949 nor year of examination 1968 qualifies as a suited instrument.

We report in Table 3.15 in appendix the corresponding estimates on having at least a *baccalauréat* (*général* or not). Indeed the events of May 1968 mainly led to a modification of the *baccalauréat* examination conditions and thus changed the probability of obtaining the *baccalauréat* and potentially pursuing higher education studies. The estimates are again small and not statistically significant, and the instrument is weak.

Overall, whichever educational outcome, the estimates are consistently insignificant. One possible explanation is that students can take the *baccalauréat* examination more than once. Indeed if they fail the first time, they have the opportunity to repeat the last grade of high school and take the examination again, until they succeed. Obtaining the *baccalauréat* at first or second try only changes the year of success, but not the fact of having it or not in the end.

Figure 3.6: Success rates at the *baccalauréat* among men for sessions 1963 to 1973 using FQP



Source: FQP 1993 and 2003.

In Figure 3.1, we observed an increase in the rate of success at the *baccalauréat* examination by nearly 20 percentage points in 1968 compared to the adjacent years.

We find similar rates of success by examination year using FQP data, as seen in Figure 3.6. However it does not correspond to the probability to ever have a *baccalauréat* when considering the possibility to take it more than once. Indeed Table 3.10 displays the success rates by year of first try at the examination, depending on whether the students took the *baccalauréat* examination for the first time, or considering the possibility to fail, repeat the last grade of high school and take it again. The rate of success indeed increased for the year 1968 for students taking the examination for the first. Nevertheless, the probability of ever obtaining the *baccalauréat* remains fairly stable over the considered period (with a slight decreasing trend at the end) and is only slightly higher in year 1968. These results should be considered with caution due to small sample size.

Table 3.10: Success rates by year of first try at the examination, for the first try or with the possibility to take it more than once

	1965	1966	1967	1968	1969	1970	1971
First try	64.67	53.59	58.44	75.45	67.60	63.12	63.64
Ever	90.00	90.85	87.66	91.02	88.27	86.52	83.42
Number of candidates	150	153	154	167	179	141	187

Source: FQP 1993 and 2003.

Sample: Men who took the *baccalauréat* examination between 1965 and 1971.

Specification: The first row corresponds to the probability of obtaining the *baccalauréat* in year t if the student took the examination for the first time in year t. The second row corresponds to the probability of ever obtaining the *baccalauréat* (in year t or later) if the student took the examination for the first time in year t.

Thus the events of May 1968 increased the likelihood to obtain the *baccalauréat* in year 1968 by about 30%, but not significantly the likelihood of ever obtaining it (which is fairly stable around 88% on average when taking it once or more). It only created an anticipation effect. A large part of individuals who benefited from the higher rate of success of 1968 would otherwise have obtained the *baccalauréat* in 1969 anyway, by repeating the last year of high school and taking the examination again. All in all the events of May 1968 did not significantly change the education level of the population affected and thus cannot be used to instrument it.

3.6 Conclusion

In this paper we investigate whether the events of May 1968 can be used as a natural experiment to instrument education. Indeed Maurin and McNally (2008) justify the use of these events by arguing that the unusually high rate of success at the *baccalauréat* in year 1968 was unique, unexpected and temporary. However not only the change in the modalities of examination led to an increase in the rate of *bacheliers*, but the *baccalauréat technologique* was also created in 1968 with a first session in 1969. Thus the opportunities offered by the education system were modified for students coming after May 1968 events.

Furthermore Maurin and McNally (2008) use birth year 1949 as an instrument. The authors do not dispose of the year individuals took their examinations and argue that individuals born in 1949 are the most likely to have taken the *baccalauréat* examination in year 1968. Nonetheless even if the median age to take the examination was indeed 19, a substantial part of students who took the examination in 1968 were likely born before or after 1949 (in particular in 1950 for students who never repeated any grade and took the examination at age 18).

To assess the validity of the natural experiment, we replicate the estimations of Maurin and McNally (2008). Using the same LFS data and strategy, we are able to accurately reproduce their results. However further investigating the first stage estimations and in particular the F-statistics, we find that not only the effect is very small but the instrument is weak. This can explain why Maurin and McNally (2008) find IV estimates much larger than their OLS ones. This replication exercise raises concerns about the validity of using being born in 1949 as an instrumental variable.

Indeed we suspect weak instrument issues as the instrument might be positively correlated with other determinants of the outcome of interest (here earnings or children's education), which would violate the exclusion restriction. One possible explanation would be that the modifications in social and moral norms caused by the events of May 1968 might have induced changes in preferences on the treated population. It would be the case if being born in 1949 and thus be treated by the events of May 1968 was correlated for example with the likelihood of being unionized, the relationship to hierarchy and/or the taste for independence (traits potentially related to labor market outcomes).

In order to dismiss the possibility that the weak instrument concern is due to small sample size, and investigate whether a larger data set would be more suitable, we compute the same estimations on Census data. The F-statistics become largely above the threshold for weak instrument. However the estimates are even smaller than those obtained from LFS data. Moreover the instrument fails at the placebo tests, as artificial alternative birth cohorts yield higher effect and F-statistics. Overall whichever the data source exploited, being born in 1949 reveals to be an unsatisfying instrument.

Thus we excluded birth year 1949, but not that an alternative instrument can be used to represent the fact of being affected by the events of May 1968. In particular the year individuals took the *baccalauréat* examination seems to be the most relevant one. We compute the first stage estimations using FQP data, as an educational calendar is provided in these surveys. We find negligible estimates associated with F-statistics hardly different from 0 and conclude that taking the *baccalauréat* examination in 1968 did not increase the likelihood to obtain the *baccalauréat* and potentially pursue higher education studies.

Since students have the opportunity to take the examination more than once, the events only created an anticipation effect. Finally, beyond any data source, sampling and estimation strategies or choice of educational outcome issues, the events of May 1968 increased the rate of success at the *baccalauréat* examination for year 1968 but had no impact on the final level of education. Thus these events cannot be used as a natural experiment.

3.A Appendix

Figure 3.7: Figure 1 in Maurin and McNally (2008)

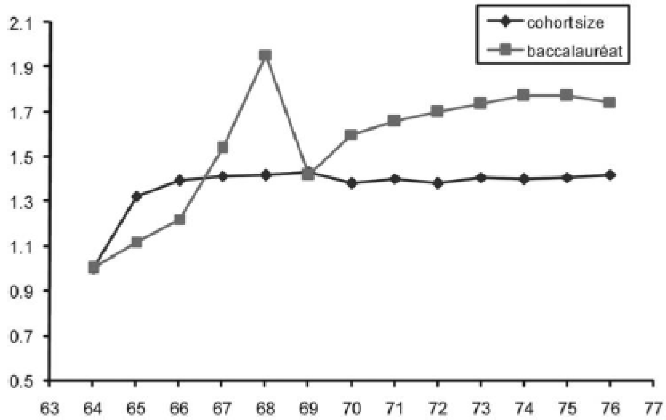


FIG. 1.—Trends in the number of *bacheliers* and in cohort size. The size of the cohort for year *t* corresponds to the number of persons born at *t* − 19 (19 is the median age of candidates). The two series are normalized to one in 1945. Source: French Ministry of Education (number of *bacheliers*) and the French Statistical Office (cohort size).

Source: Maurin and McNally (2008).

Table 3.11: *Baccalauréat* examination from 1945 to 1953 (men and women)

Birth year	Cohort size	Year of examination	*Baccalauréat* Taking the examination	*Baccalauréat* Passing the examination	*Baccalauréat* Success rate	*Baccalauréat général* Taking the examination	*Baccalauréat général* Passing the examination	*Baccalauréat général* Success rate
1945	645,899	1964	138,430	86,729	62.7	138,430	86,729	62.7
1946	843,904	1965	159,186	96,924	60.9	159,186	96,924	60.9
1947	870,472	1966	212,420	105,839	49.8	212,420	105,839	49.8
1948	870,836	1967	223,410	133,257	59.6	223,410	133,257	59.6
1949	872,661	1968	208,460	169,422	81.3	208,460	169,422	81.3
1950	862,310	1969	207,682	137,015	66.0	181,466	122,673	67.6
1951	826,722	1970	249,120	167,307	67.2	200,722	138,707	69.1
1952	822,204	1971	272,009	176,766	65.0	217,298	143,729	66.1
1953	804,696	1972	282,263	184,196	65.3	221,453	147,352	66.5

Source: French Statistical Office INSEE (cohort size) and French Ministry of Education (examinations data).
Following Maurin and McNally (2008), the birth year t corresponds to the year of examination $t + 19$
(19 is the median age for the candidates).

Table 3.12: Additional descriptive statistics using LFS, Census and FQP

	LFS				**Census**		**FQP**	
	Male wage earners		All men		All men		All men	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Education outcomes								
Baccalauréat at least	0.278	(0.445)	0.272	(0.442)	0.301	(0.459)	0.238	(0.426)
Baccalauréat général at least	0.263	(0.443)	0.220	(0.414)	0.232	(0.422)	0.197	(0.397)
Observations	26,293		36,629		2,488,383		4,828	

Source: LFS 1990, 1993, 1996 and 1999, Census 1999 and FQP 1993 and 2003.
Sample: Men born between 1946 and 1952.

As reported in Table 3.12, 27.8% of the LFS sample of male wage earners have at least a *baccalauréat* and 26.3% have at least a *baccalauréat général*. The corresponding proportions for all men are 27.2% and 22.0% in the LFS sample, 30.1% and 23.2% in the Census sample, and 23.8% and 19.7% in the FQP sample.

Table 3.13: First stage estimations: Effect of $May68$ on alternative education outcomes using LFS and Census

	LFS				Census	
	Male wage earners		All men		All men	
	Baccalauréat at least	*Baccalauréat général* at least	*Baccalauréat* at least	*Baccalauréat général* at least	*Baccalauréat* at least	*Baccalauréat général* at least
	(1)	(2)	(3)	(4)	(5)	(6)
$May68$	0.029***	0.029***	0.021***	0.017**	0.008***	0.008***
	(0.001)	(0.000)	(0.007)	(0.007)	(0.001)	(0.001)
Birth year	0.007***	0.004**	0.006***	0.004**	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
F-stat	10.79	12.02	7.59	6.09	65.64	85.34
R-squared	0.001	0.002	0.001	0.001	0.000	0.000
Observations	11,145	11,145	15,433	15,433	1,056,384	1,056,384

Source: LFS 1990, 1993, 1996 and 1999, and Census 1999.
Sample: Male wage earners/men born in 1946, 1949 and 1952.
Specification: $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or 1952.
Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 3.13 shows that being born in 1949 (versus 1946 or 1952) increases the likelihood of having at least a *baccalauréat* (of any kind or *général*) by about 3 percentage points for the LFS sample – when considering wage earners – which corresponds to a probability of around 31% instead of 28% for all kinds of *baccalauréats* and around 29% instead of 26% for the *baccalauréat général*. When considering all men from LFS data, the increase is about 2%, leading to a probability of 29% instead of 27% for all kinds of *baccalauréats* and around 24% instead of 22% for the *baccalauréat général*. Corresponding rates for the Census sample are 31% instead of 30% and 24% instead of 23% respectively, as the increase is only of 0.8 percentage points. The results obtained from LFS data are about 2 to 3 times larger than the ones obtained from Census data.

Table 3.14: Evaluation of the returns to education using alternative education outcomes on LFS

	Log wage			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Baccalauréat général at least	0.511***	0.755**		
	(0.010)	(0.318)		
Baccalauréat at least			0.503***	0.746**
			(0.009)	(0.311)
Birth year	0.010***	0.010***	0.009***	0.008***
	(0.002)	(0.003)	(0.002)	(0.003)
R-squared	0.21		0.23	
Observations	11,145	11,145	11,145	11,145

Source: LFS 1990, 1993, 1996 and 1999.

Sample: Male wage earners born in 1946, 1949 and 1952.

Specification: $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or 1952.

Table 3.15: First stage estimations: Effect of $May68$ on alternative education outcomes using FQP, with either birth year or year of examination as an instrumental variable

	Birth year 1949		**Year of examination**	
	Baccalauréat at least	*Baccalauréat général* at least	*Baccalauréat* at least	*Baccalauréat général* at least
	(1)	(2)	(3)	(4)
$May68$	0.012	0.014	0.022	0.037
	(0.020)	(0.019)	(0.025)	(0.031)
Birth year	0.002	-0.003	-0.010**	-0.027***
	(0.004)	(0.004)	(0.005)	(0.006)
F-stat	0.37	0.57	0.74	1.39
R-squared	0.000	0.001	0.005	0.019
Observations	2,015	2,015	1,130	1,130

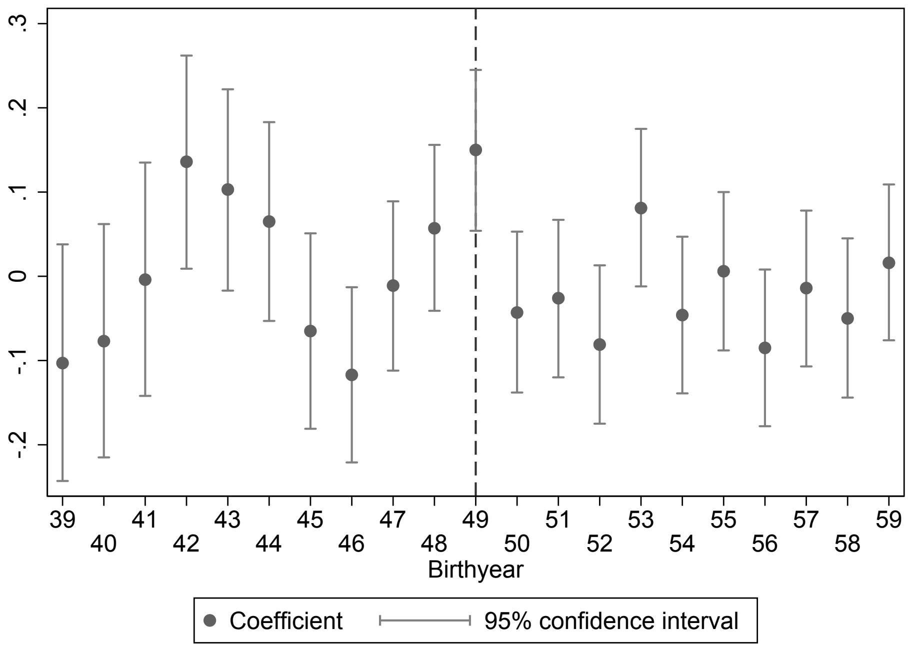
Source: FQP 1993 and 2003.

Sample: For (1) and (2), individuals born in 1946, 1949 and 1952; for (3) and (4), individuals born between 1946 and 1952 who took the *baccalauréat* examination.

Specification: For (1) and (2), $May68$ equals 1 if the individual is born in 1949, 0 if the individual is born in 1946 or1952; for (3) and (4), $May68$ equals 1 if the individual took the *baccalauréat* examination in 1968.

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Figure 3.8: Robustness checks for men born from 1939 to 1959, using LFS data, outcome: Years of higher education



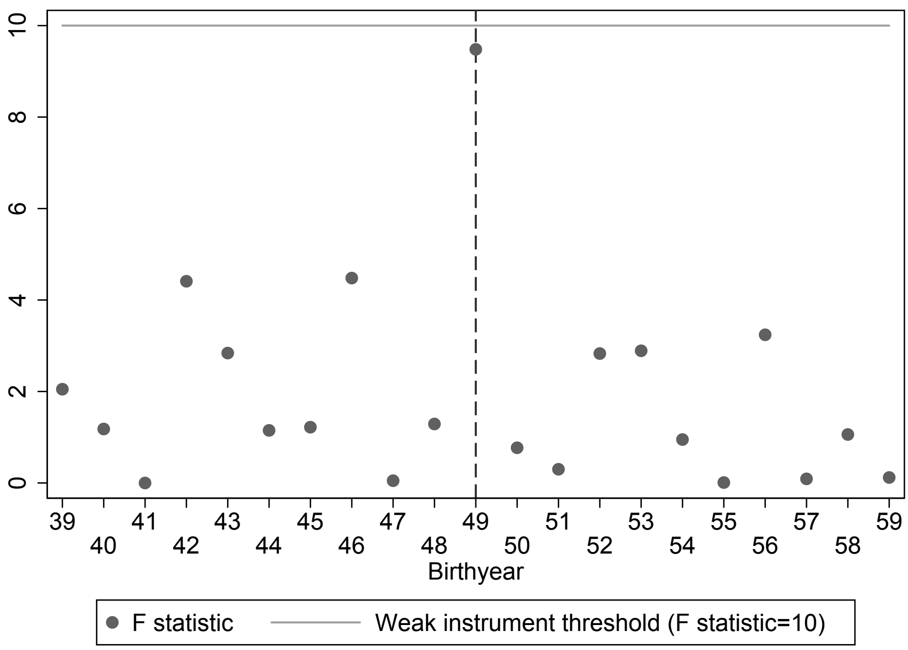


Figure 3.9: Robustness checks for men born from 1939 to 1959, using Census data, outcome: Years of higher education

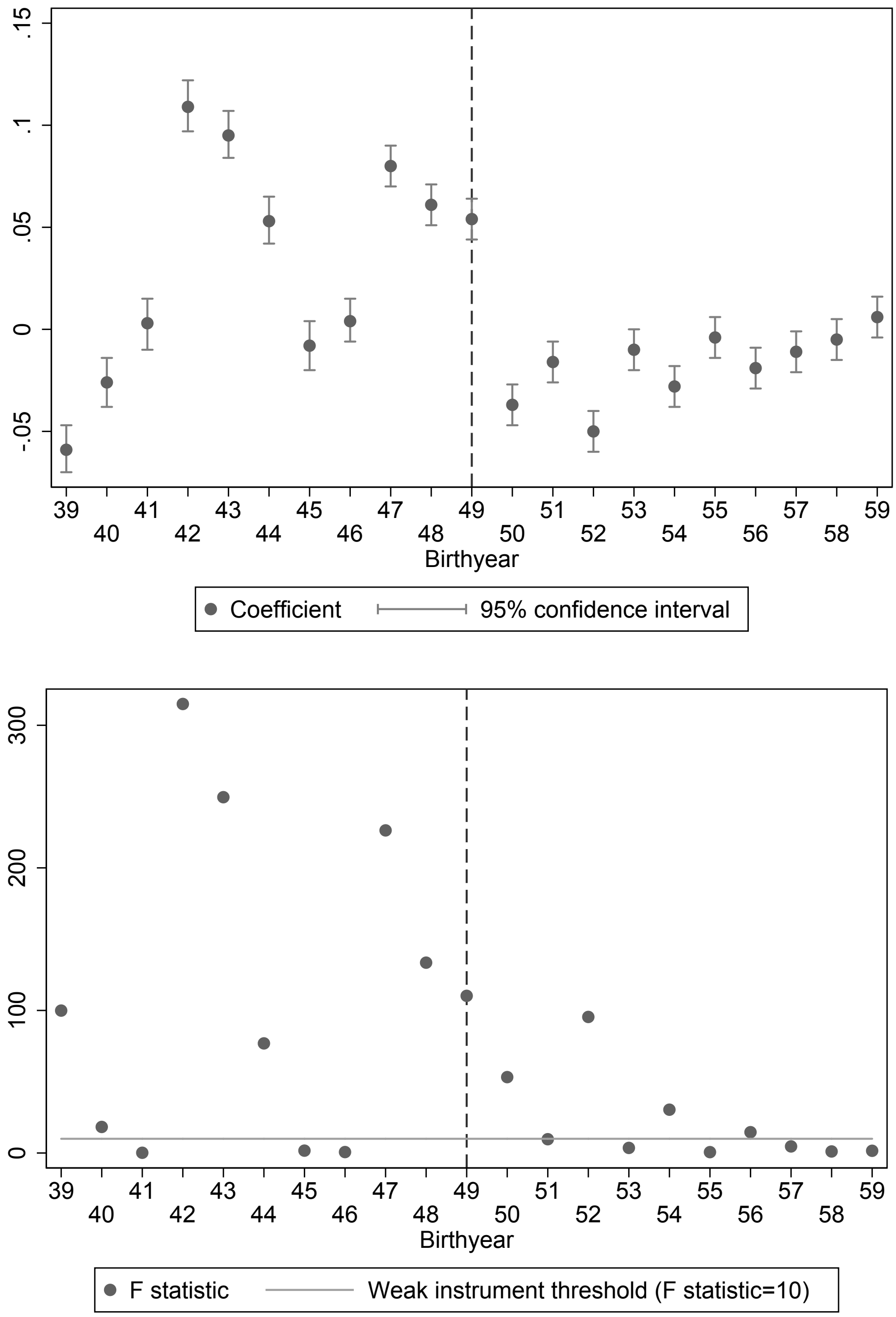
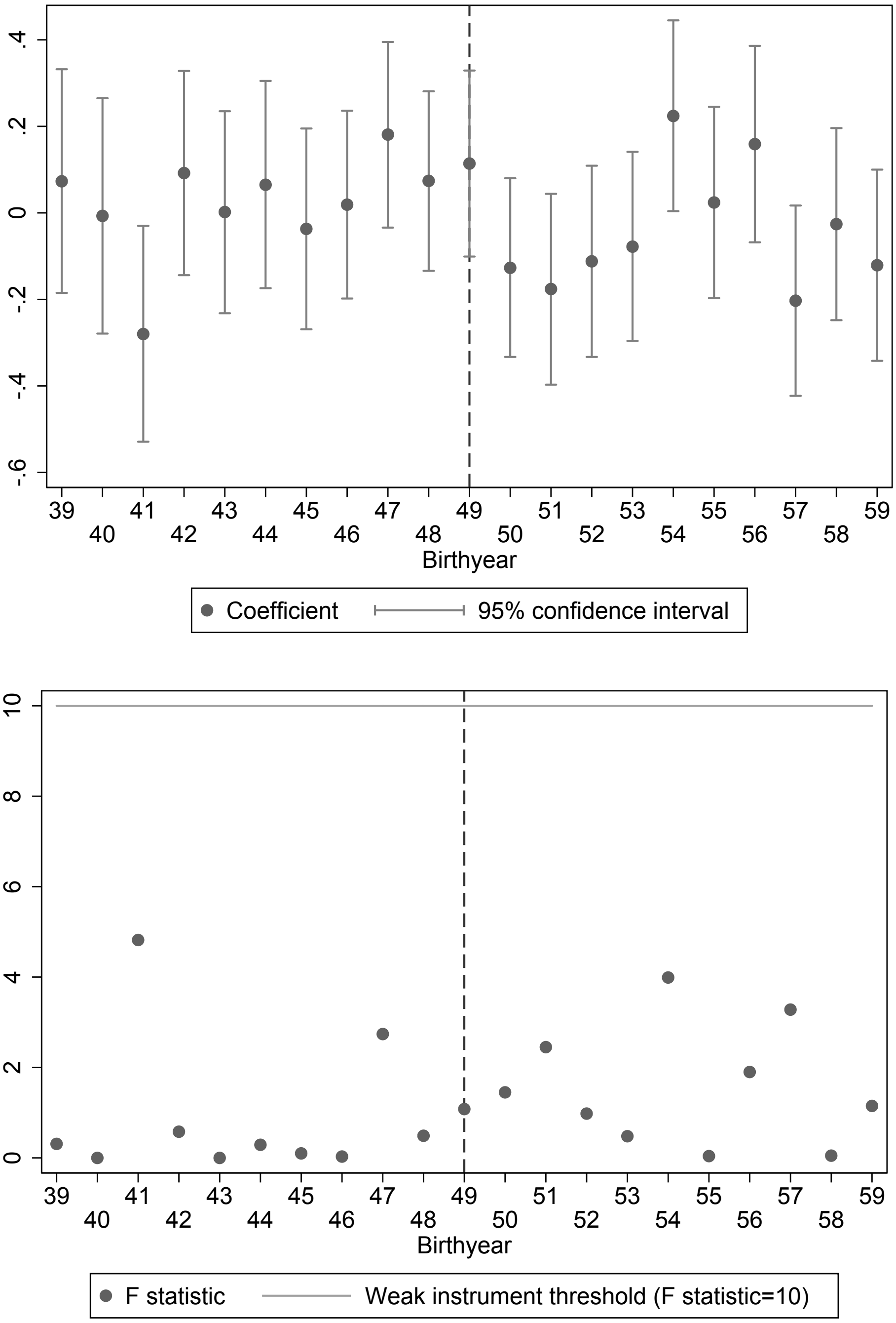


Figure 3.10: Robustness checks for men born from 1939 to 1959, using FQP data, outcome: Years of higher education



Conclusion

This thesis investigates intergenerational mobility in three European countries: Germany, France and Sweden. The aim is to evaluate the extent and the determinants of the transmission of socioeconomic inequality from one generation to the next. This constitutes an important question as it highlights whether individuals only reproduce or can depart from the socioeconomic status of their parents, in other words whether their family background plays a determinant role in their future success in life, and thus whether they benefit from equality of opportunity.

The first and most intuitive approach we implement is to estimate the log-linear intergenerational economic regression linking father's to son's earnings to obtain an estimation of the intergenerational elasticity, in Germany. This measure of intergenerational transmission of inequalities can suffer from severe misestimation if biases arising from measurement error of both generations' permanent earnings are not correctly handled, which is why these biases are to be carefully assessed and treated. One drawback of this approach is that the intergenerational elasticity only focuses on a single family characteristic – here father's earnings – and thus does not account for all channels through which children's socioeconomic success can be influenced by family background.

In the first chapter we carefully examine the literature estimating the intergenerational earnings transmission in Germany and reveal uncertainty about its extent. Indeed as mentioned, the estimation can be severely biased by measurement error in the permanent earnings of both generations. Whether and how these biases are treated yields substantially different estimates. To evaluate the magnitude of the biases and identify the best way to handle them, we first depict the association between current and lifetime earnings over the life-cycle, using data from the SOEP, and confirm the international evidence suggesting to observe earnings of sons at least in their thirties and fathers before their fifties. Then we estimate the intergenerational elasticity at 0.323 using our main sampling strategy, restricting the analysis to sons aged 35 to 42 and fathers aged 30 to 55 and presenting at least five available earnings observations, which is in line with the most recent German literature.

Then, we implement a different strategy through the estimation of sibling correlations in France and in Sweden, as they constitute a broader indicator of the influence of family and community on children's adult attainments. Moreover a single economic outcome is considered in the first chapter, whereas the second chapter investigates three different measures of socioeconomic success in France: education, profession and earnings. A comparative study of the extent of the brother earnings correlation in France and Sweden is then conducted, both to assess the relative levels of mobility of these two countries and for methodological concerns, to evaluate the impact of using predicted earnings instead of actual permanent earnings on the estimation, when extensive enough data are not available.

In the second chapter, we estimate sibling correlations in predicted variables of education, occupation and earnings, in France, using FQP data. Our estimates amount respectively 0.522, 0.336 and 0.459 for the three socioeconomic outcomes, and reveal that a substantial part of the transmission of inequalities had not been taken into account by the estimation of the intergenerational elasticity. We also investigate the influence of several family characteristics – such as sibship's size and composition, as well as parental education and occupation – on the level of transmission of the socioeconomic status. We show in particular that sibling correlations are higher for same-sex siblings than between brothers and sisters, in larger than in smaller families and when parents present lower levels of education and occupation. We also estimate brother earnings correlations for France and Sweden, to assess the relative extents of mobility in the two countries, as well as to evaluate the impact of using predicted instead of actual permanent earnings measures, when the corresponding data are not available, as it is the case for France. For the case of Sweden, we estimate the brother correlation in permanent earnings at 0.265, and our results suggest that the corresponding correlation amounts around 0.4 in France, higher than in Nordic countries, smaller than in the United States and close to other Western European countries.

Lastly one essential aspect of the analysis of intergenerational mobility lies not only in the investigation of its extent, but also of the transmission mechanisms involved, and in particular the causal link between parental and children's education or income. Indeed if public policies do not – and should not – have any impact on inequalities due to individual choices, they however should attempt to equalize life prospects if inequalities are directly due to parental success, as the socioeconomic status of the family in which an individual is born is a matter of chance. The method considered here is the one exploiting natural experiments. Indeed, if a particular event or reform which induced an exogenous modification of the education level of a specific population – uncorrelated with any individual characteristic – can be identified, it can be used to assess the causal effect of the educational change on the next generation. Therefore, we investigate the possibility to use the events of May 1968 in France as a natural experiment.

In the third chapter, we thus investigate the possibility to use these events to evaluate the causal link between the education levels of parents and children. Indeed the rate of success at the examination ending high school and giving access to higher education in France – the *baccalauréat* – was higher in 1968 than for the adjacent years, due to a modification of the modalities of examination this year, following the revolt. It can be argued that this constitutes an exogenous increase in the education level of the students who took this examination in year 1968 and thus that this particular situation can be used as a natural experiment. However our estimations, conducted on data from the LFS, the Census and the FQP surveys, are consistently negligible and/or fail at robustness checks, for both instruments investigated: being born in 1949 and thus being 19 in 1968, and having taken the *baccalauréat* examination in 1968. We conclude that the events of May 1968 increased the rate of success at the *baccalauréat* examination in 1968, but not the final level of education of the affected population, as students had the opportunity to take the examination more than once by repeating a grade. Thus the only effect was for more students to obtain the *baccalauréat* at their first try in 1968, and the events of May 1968 do not constitute a favorable framework for a natural experiment strategy to instrument education.

Finally, the main contributions of this thesis studying the intergenerational transmission of inequalities in Europe are the following. We implement different strategies, highlight the issues that can arise due to incorrect or insufficient treatment of biases, or to data limitations, and suggest methods to handle them and to check the validity of the estimations. In the first chapter, we reassess the extent of intergenerational transmission of earnings in Germany. We carefully address the biases issues notably by estimating for the first time their extent using SOEP data, and confirm the most recent estimations of the intergenerational elasticity slightly above 0.3. In the second chapter, we fill a gap in the literature by providing a first estimation of sibling correlations for France, in terms of education, occupation and earnings, and conclude that a substantial part of the transmission had not been taken into account by estimations of intergenerational elasticities. To do so, we have to predict continuous measures of success for two siblings, due to a lack of extensive enough data in France. Therefore, we also compute comparable brother correlations for France and Sweden to be able to assess the impact of using predicted variables on the estimation results. We conclude that the actual brother correlation in earnings lies around 0.4 in France. In the third chapter, we study the estimation of the causal link between parents' and children's education. We consider the use of the events of May 1968 in France as a natural experiment, but show that they do not constitute a suitable framework to instrument education. We conclude that one cannot exploit a supposed exogenous increase in the education level following these events.

Bibliography

AAKVIK, A., K. SALVANES, AND K. VAAGE (2010): "Measuring heterogeneity in the returns to education using an education reform," *European Economic Review*, 54, 483–500.

AARONSON, D. AND B. MAZUMDER (2008): "Intergenerational economic mobility in the United States, 1940 to 2000," *The Journal of Human Resources*, 43, 139–172.

ACEMOGLU, D. AND J. ANGRIST (1999): "How large are the social returns to education? Evidence from compulsory schooling laws," National Bureau of Economic Research.

ALTONJI, J. AND T. DUNN (1991): "Relationships among the family incomes and labor market outcomes of relatives," in *Research in Labor Economics*, ed. by R. Ehrenberg, CT: JAI Press, vol. 12, 269–310.

ANGRIST, J. AND A. KRUEGER (1991): "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

——— (1992): "The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples," *Journal of the American Statistical Association*, 87, 328–336.

——— (1999): "Empirical strategies in labor economics," *Handbook of labor economics*, 3, 1277–1366.

ANTONOVICS, K. AND A. GOLDBERGER (2005): "Does increasing women's schooling raise the schooling of the next generation? Comment," *The American Economic Review*, 95, 1738–1744.

ASHENFELTER, O. AND A. KRUEGER (1994): "Estimates of the economic return to schooling from a new sample of twins," *The American Economic Review*, 84, 1157–1173.

Becker, G. and N. Tomes (1976): "Child endowments and the quantity and quality of children," *Journal of political Economy*, 84, S143–S162.

——— (1979): "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy*, 872, 1153–1189.

——— (1986): "Human capital and the rise and fall of families," *Journal of Labor Economics*, 4, 1–39.

Behrman, J., R. Pollak, and P. Taubman (1982): "Parental preferences and provision for progeny," *Journal of Political Economy*, 90, 52–73.

——— (1986): "Do parents favor boys?" *International Economic Review*, 33–54.

Behrman, J. and M. Rosenzweig (2002): "Does increasing women's schooling raise the schooling of the next generation?" *The American Economic Review*, 92, 323–334.

——— (2005): "Does increasing women's schooling raise the schooling of the next generation? Reply," *The American Economic Review*, 95, 1745–1751.

Behrman, J., M. Rosenzweig, and P. Taubman (1994): "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment," *Journal of political economy*, 102, 1131–1174.

Behrman, J. and P. Taubman (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*, 144–151.

——— (1986): "Birth order, schooling, and earnings," *Journal of Labor Economics*, 4, S121–S145.

Bingley, P. and L. Cappellari (2013): "Correlations of brothers earnings and intergenerational transmission," Dipartimento di Economia e Finanza, Università Cattolica.

Bingley, P., K. Christensen, and V. Jensen (2009): "Parental schooling and child development: Learning from twin parents," The Danish National Centre for Social Research, Social Policy and Welfare Services.

Björklund, A., T. Eriksson, M. Jäntti, O. Raaum, and E. Österbacka (2002): "Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States," *Journal of Population Economics*, 15, 757–772.

Björklund, A. and M. Jäntti (1997): "Intergenerational income mobility in Sweden compared to the United States," *The American Economic Review*, 87, 1009–1018.

——— (2000): "Intergenerational mobility of socio-economic status in comparative perspective," *Nordic Journal of Political Economy*, 26, 3–32.

——— (2009): "Intergenerational income mobility and the role of family background," in *Oxford Handbook of Economic Inequality*, ed. by W. Salverda, B. Nolan, and T. Smeeding, Oxford: Oxford University Press, chap. 20, 491–521.

——— (2012): "How Important is Family Background for Labor-Economic Outcomes?" *Labour Economics*, 19, 465–74.

Björklund, A., M. Jäntti, and M. Lindquist (2009): "Family background and income during the rise of the welfare state: Brother correlations in income for Swedish men born 1932-1968," *Journal of Public Economics*, 93, 671–80.

Björklund, A., L. Lindahl, and M. Lindquist (2010): "What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents," *The BE Journal of Economic Analysis & Policy*, 10, 102.

Björklund, A., M. Lindahl, and E. Plug (2004): "Intergenerational effects in Sweden: What can we learn from adoption data?" IZA Discussion Paper.

——— (2006): "The origins of intergenerational associations: Lessons from Swedish adoption data," *The Quarterly Journal of Economics*, 999–1028.

Björklund, A. and K. Salvanes (2010): "Education and family background: Mechanisms and policies," IZA Discussion Paper.

Black, S. and P. Devereux (2011): "Recent developments in intergenerational mobility," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Vol. 4 Part B, 1487–1541.

Black, S., P. Devereux, and K. Salvanes (2005): "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital," *The American Economic Review*, 95, 437–449.

Blackburn, M. and D. Neumark (1993): "Are OLS estimates of the return to schooling biased downward? Another look," National Bureau of Economic Research.

Blanden, J. (2005): *Essays on intergenerational mobility and its variation over time, place and family structure*, PhD Thesis, University of London.

——— (2013): "Cross-country rankings in intergenerational mobility: A comparison of approaches from economics and sociology," *Journal of Economic Surveys*, 27, 38–73.

Blanden, J., A. Goodman, P. Gregg, and S. Machin (2004): "Changes in intergenerational mobility in Britain," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 122–146.

Böhlmark, A. and M. Lindquist (2006): "Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden," *Journal of Labor Economics*, 24, 879–896.

Bound, J. and D. Jaeger (1996): "On the validity of season of birth as an instrument in wage equations: A comment on Angrist & Krueger's "Does compulsory school attendance affect schooling and earnings?"," National Bureau of Economic Research.

——— (2000): "Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?" in *Research in labor economics*, Emerald Group Publishing Limited, 83–108.

Bound, J., D. Jaeger, and R. Baker (1993): "The cure can be worse than the disease: A cautionary tale regarding instrumental variables," National Bureau of Economic Research.

——— (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American statistical association*, 90, 443–450.

Boutchénik, B., C. Coron, S. Grobon, C. Goffette, and L.-A. Vallet (2015): "Quantifying the total influence of family of origin on the schooling and professional careers of individuals," *Economie & Statistique*, 477, 5–23.

Bratberg, E., Ø. Nilsen, and K. Vaage (2003): "Assessing changes in intergenerational earnings mobility," IZA Discussion Paper.

Bratsberg, B., K. Røed, O. Raaum, R. Naylor, M. Jäntti, T. Eriksson, and E. Österbacka (2007): "Nonlinearities in intergenerational earnings mobility: consequences for cross-country comparisons," *The Economic Journal*, 117.

Breen, R. (2004): *Social mobility in Europe*, Oxford University Press.

Brenner, J. (2010): "Life-cycle variations in the association between current and lifetime earnings: Evidence for German natives and guest workers," *Labour Economics*, 17, 392–406.

Butcher, K. and A. Case (1994): "The effect of sibling sex composition on women's education and earnings," *The Quarterly Journal of Economics*, 109, 531–563.

Cappellari, L. and C. Lucifora (2009): "The Bologna Process and college enrollment decisions," *Labour Economics*, 16, 638–647.

Card, D. (1999): "The causal effect of education on earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, vol. 3, 1801–1863.

——— (2001): "Estimating the return to schooling: Progress on some persistent econometric problems," *Econometrica*, 69, 1127–1160.

Carneiro, P., C. Meghir, and M. Parey (2013): "Maternal education, home environments, and the development of children and adolescents," *Journal of the European Economic Association*, 11, 123–160.

Chadwick, L. and G. Solon (2002): "Intergenerational income mobility among daughters," *The American Economic Review*, 92, 335–344.

Chambaz, C., E. Maurin, and C. Torelli (1998): "L'évaluation sociale des professions en France. Construction et analyse d'une échelle des professions." *Revue française de sociologie*, 39, 177–226.

Chevalier, A. (2004): "Parental education and child's education: A natural experiment," IZA Discussion Paper.

Conley, D. (2009): *The pecking order: A bold new look at how family and society determine who we become*, Vintage.

Conley, D. and R. Glauber (2008): "All in the family? Family composition, resources, and sibling similarity in socioeconomic status," *Research in Social Stratification and Mobility*, 26, 297–306.

Corak, M. (2004): "Generational income mobility in North America and Europe: an introduction," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 1–37.

——— (2006): "Do poor children become poor adults? Lessons from a cross-country comparison of generational earnings mobility," in *Dynamics of inequality and poverty*, Emerald Group Publishing Limited, vol. 13, 143–188.

——— (2013): "Income inequality, equality of opportunity, and intergenerational mobility," *The Journal of Economic Perspectives*, 27, 79–102.

Corak, M. and A. Heisz (1999): "The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data," *Journal of Human Resources*, 504–533.

Corcoran, M., R. Gordon, D. Laren, and G. Solon (1990): "Effects of family and community background on economic status," *The American Economic Review*, 80, 362–366.

Corcoran, M., C. Jencks, and M. Olneck (1976): "The effects of family background on earnings," *The American Economic Review*, 66, 430–435.

Couch, K. and T. Dunn (1997): "Intergenerational correlations in labor market status: A comparison of the United States and Germany," *Journal of Human Resources*, 32, 210–232.

——— (1999): "Intergenerational correlations in earnings in three countries: The United Kingdom, Germany and the United States," *Vierteljahrshefte zur Wirtschaftsforschung*, 68, 290–296.

Couch, K. and D. Lillard (2004): "Nonlinear patterns of intergenerational mobility in Germany and the United States," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 190–206.

Currie, J. and E. Moretti (2003): "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *The Quarterly Journal of Economics*, 118, 1495–1532.

Dearden, L., S. Machin, and H. Reed (1997): "Intergenerational mobility in Britain," *The Economic Journal*, 107, 47–66.

Drasgow, F. (1988): "Polychoric and polyserial correlations," *Encyclopedia of statistical sciences*.

Duflo, E. (2000): "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," National Bureau of Economic Research.

Eide, E. and M. Showalter (1999): "Factors affecting the transmission of earnings across generations: A quantile regression approach," *Journal of Human Resources*, 253–267.

Eisenhauer, P. and F. Pfeiffer (2008): "Assessing intergenerational earnings persistence among German workers," *Journal for Labour Market Research*, 41, 119–137.

Erikson, R. and J. Goldthorpe (1992): *The constant flux: A study of class mobility in industrial societies*, Oxford University Press, USA.

——— (2002): "Intergenerational inequality: A sociological perspective," *The Journal of Economic Perspectives*, 16, 31–44.

——— (2010): "Has social mobility in Britain decreased? Reconciling divergent findings on income and class mobility," *The British journal of sociology*, 61, 211–230.

Ermisch, J. and M. Francesconi (2004): "Intergenerational mobility in Britain: New evidence from the British Household Panel Survey," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 147–189.

Ermisch, J., M. Francesconi, and T. Siedler (2006): "Intergenerational mobility and marital sorting," *The Economic Journal*, 116, 659–679.

Fertig, A. (2003): "Trends in intergenerational earnings mobility in the US," *Journal of Income Distribution*, 12.

Fisher, R. (1915): "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, 10, 507–521.

Frick, J., S. Jenkins, D. Lillard, O. Lipps, and M. Wooden (2007): "The cross-national equivalent file (CNEF) and its member country household panel studies," *Journal of Applied Social Science Studies*, 127, 627–654.

Friedman, M. (1957): "The permanent income hypothesis," in *A theory of the consumption function*, Princeton University Press, 20–37.

GRAWE, N. (2004): "Intergenerational mobility for whom? The experience of high- and low-earning sons in international perspective," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 58–89.

——— (2006): "Lifecycle bias in estimates of intergenerational earnings persistence," *Labour Ecnonomics*, 13, 551–570.

GRENET, J. (2013): "Is extending compulsory schooling alone enough to raise earnings? Evidence from French and British compulsory schooling laws," *The Scandinavian Journal of Economics*, 115, 176–210.

GURGAND, M. AND E. MAURIN (2007): "A large scale experiment: wages and educational expansion in France," PSE Working Papers No. 2007-21.

HAEGELAND, T., L. KIRKEBØEN, O. RAAUM, AND K. SALVANES (2010): "Why children of college graduates outperform their schoolmates: A study of cousins and adoptees," NHH Dept. of Economics Discussion Paper.

HAIDER, S. AND G. SOLON (2006): "Life-cycle variation in the association between current and lifetime earnings," *The American Economic Review*, 96, 1308–1320.

HERTZ, T. (2007): "Trends in the intergenerational elasticity of family income in the United States," *Industrial Relations: A Journal of Economy and Society*, 46, 22–50.

HIRVONEN, L. (2008): "Intergenerational earnings mobility among daughters and sons: Evidence from Sweden and a comparison with the United States," *American Journal of Economics and Sociology*, 67, 777–826.

HOLMLUND, H. (2006): "Intergenerational mobility and assortative mating. Effects of an educational reform," Swedish Institute for Social Research (SOFI) Working Paper 4/2006.

HOLMLUND, H., M. LINDAHL, AND E. PLUG (2011): "The causal effect of parents' schooling on children's schooling: A comparison of estimation methods," *Journal of Economic Literature*, 49, 615–651.

ICHINO, A. AND R. WINTER-EBMER (2004): "The long-run educational cost of World War II," *Journal of Labor Economics*, 22, 57–87.

Jäntti, M., B. Bratsberg, K. Røed, O. Raaum, R. Naylor, E. Österbacka, A. Björklund, and T. Eriksson (2006): "American exceptionalism in a new light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States," IZA Discussion Paper.

Jenkins, S. (1987): "Snapshots vs movies: 'lifecycle biases' and the estimation of intergenerational earnings inheritance," *European Economic Review*, 31, 1149–1158.

Knigge, A. (2015): *Sources of sibling similarity: Status attainment in the Netherlands during modernization*, ICS-dissertation, Utrecht.

Lee, C. and G. Solon (2009): "Trends in intergenerational income mobility," *Review of Economics and Statistics*, 91, 766–772.

Lefranc, A. (2011): "Educational expansion, earnings compression and changes in intergenerational economic mobility: Evidence from French cohorts, 1931-1976," THEMA Working Papers.

Lefranc, A. and A. Trannoy (2005): "Intergenerational earnings mobility in France: Is France more mobile than the US?" *Annales d'économie et de statistique*, 78.

Leigh, A. and C. Ryan (2008): "Estimating returns to education using different natural experiment techniques," *Economics of Education Review*, 27, 149–160.

Levine, D. and B. Mazumder (2002): "Choosing the right parents: Changes in the intergenerational transmission of inequality between 1980 and the early 1990s," Federal Reserve Bank of Chicago.

——— (2007): "The growing importance of family: Evidence from brothers' earnings," *Industrial Relations: A Journal of Economy and Society*, 46, 7–21.

Lillard, D. (2001): "Cross-national estimates of the intergenerational mobility in earnings," *Vierteljahrshefte zur Wirtschaftsforschung*, 70, 51–58.

Maurin, E. and S. McNally (2008): "Vive la révolution! Long-term educational returns of 1968 to the angry students," *Journal of Labor Economics*, 26, 1–33.

Mayer, S. and L. Lopoo (2004): "Trends in the intergenerational transmission of economic status," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 90–121.

——— (2005): "Has the intergenerational transmission of economic status changed?" *Journal of Human Resources*, 40, 169–185.

Mazumder, B. (2005): "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data," *Review of Economics ans Statistics*, 87, 235–255.

——— (2008): "Sibling similarities and economic inequality in the US," *Journal of Population Economics*, 21, 685–701.

——— (2011): "Family and community influences on health and socioeconomic status: sibling correlations over the life course," *The BE Journal of Economic Analysis & Policy*, 11.

Meghir, C. and M. Palme (2005): "Educational reform, ability, and family background," *The American Economic Review*, 95, 414–424.

Mocetti, S. (2007): "Intergenerational earnings mobility in Italy," *The BE Journal of Economic Analysis & Policy*, 7.

Mulligan, C. (1997): *Parental priorities and economic inequality*, University of Chicago Press.

Nicoletti, C. and J. Ermisch (2008): "Intergenerational earnings mobility: Changes across cohorts in Britain," *The BE Journal of Economic Analysis & Policy*, 7.

Nybom, M. and J. Stuhler (2011): "Heterogeneous income profiles and life-cycle bias in intergenerational mobility estimation," Discussion paper series, Forschungsinstitut zur Zukunft der Arbeit.

Oettinger, G. (1999): "Does the sibling correlation in economic status vary across families and sibling pairs?" Department of Economics, University of Texas.

Oreopoulos, P. (2003): "The long-run consequences of living in a poor neighborhood," *The Quarterly Journal of Economics*, 118, 1533–1575.

——— (2006): "Estimating average and local average treatment effects of education when compulsory schooling laws really matter," *The American Economic Review*, 96, 152–175.

Oreopoulos, P., M. Page, and A. Stevens (2006): "The intergenerational effects of compulsory schooling," *Journal of Labor Economics*, 24, 729–760.

Page, M. and G. Solon (2003a): "Correlations between brothers and neighboring boys in their adult earnings: The importance of being urban," *Journal of Labor Economics*, 21, 831–855.

——— (2003b): "Correlations between sisters and neighboring girls in their subsequent income as adults," *Journal of Applied Econometrics*, 18, 545–562.

Pekkala, S. and R. Lucas (2007): "Differences across cohorts in Finnish intergenerational income mobility," *Industrial Relations: A Journal of Economy and Society*, 46, 81–111.

Peters, E. (1992): "Patterns of intergenerational mobility in income and earnings," *The Review of Economics and Statistics*, 456–466.

Piraino, P. (2007): "Comparable estimates of intergenerational income mobility in Italy," *The BE Journal of Economic Analysis & Policy*, 7.

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

Raaum, O., B. Bratsberg, K. Røed, E. Österbacka, T. Eriksson, M. Jäntti, and R. Naylor (2007): "Marital sorting, household labor supply, and intergenerational earnings mobility across countries," *The BE Journal of Economic Analysis & Policy*, 7, 1–46.

Raaum, O., K. G. Salvanes, and E. Sørensen (2006): "The neighbourhood is not what it used to be," *The Economic Journal*, 116, 200–222.

Reville, R. (1995): "Intertemporal and life cycle variation in measured intergenerational earnings mobility," RAND.

Roemer, J. (1998): *Equality of opportunity*, Harvard University Press, Cambridge, MA.

——— (2004): "Equal opportunity and intergenerational mobility: Going beyond intergenerational income transition matrices," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 48–57.

——— (2012): "What is the justification of studying intergenerational mobility of socioeconomic status?" in *From Parents to Children: The Intergenerational Transmission of Advantage*, Russell Sage Foundation, New York, 482–488.

Sacerdote, B. (2000): "The nature and nurture of economic outcomes," National Bureau of Economic Research.

——— (2007): "How large are the effects from changes in family environment? A study of Korean American adoptees," *The Quarterly Journal of Economics*, 119–157.

Schnitzlein, D. (2009): "Structure and extent of intergenerational mobility in Germany," *Journal of Economics and Statistics*, 229, 450–466.

——— (2014): "How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark," *Journal of Population Economics*, 27, 69–89.

——— (2015): "A new look at intergenerational mobility in Germany compared to the US," *Review of Income and Wealth*.

Sekhon, J. S. and R. Titiunik (2012): "When natural experiments are neither natural nor experiments," *American Political Science Review*, 106, 35–57.

Sieben, I. (2001): *Sibling similarities and social stratification: The impact of family background across countries and cohorts*, ICS-dissertation, Nijmegen.

Solon, G. (1989): "Biases in the estimation of intergenerational earnings correlations," *Review of Economics and Statistics*, 71, 172–174.

——— (1992): "Intergenerational income mobility in the United States," *The American Economic Review*, 82, 393–408.

——— (1999): "Intergenerational mobility in the labor market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Vol. 3 Part A, 1761–1800.

——— (2002): "Cross-country differences in intergenerational earnings mobility," *Journal of Economic Perspectives*, 16, 59–66.

——— (2004): "A model of intergenerational mobility variation over time and place," in *Generational income mobility in North America and Europe*, ed. by M. Corak, Cambridge University Press, 38–47.

Solon, G., M. Corcoran, R. Gordon, and D. Laren (1991): "A longitudinal analysis of sibling correlations in economic status," *Journal of Human resources*, 509–534.

Solon, G., M. Page, and G. Duncan (2000): "Correlations between neighboring children in their subsequent educational attainment," *Review of Economics and Statistics*, 82, 383–392.

Staiger, D. and J. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

Torche, F. (2015): "Analyses of intergenerational mobility: An interdisciplinary review," *The ANNALS of the American Academy of Political and Social Science*, 657, 37–62.

Treiman, D. and H. Ganzeboom (1990): "Cross-national comparative status attainment research," *Research in Social Stratification and Mobility*, 9, 105–127.

Vogel, T. (2006): "Reassessing intergenerational mobility in Germany and the United States: the impact of differences in lifecycle earnings patterns," SFB 649 Discussion Papers.

Wagner, G., J. Frick, and J. Schupp (2007): "The German Socio-economic Panel Study (SOEP)-scope, evolution and enhancements," *Journal of Applied Social Science Studies*, 127, 139–169.

Zimmerman, D. (1992): "Regression toward mediocrity in economic stature," *The American Economic Review*, 82, 409–429.

Intergenerational mobility:
An international estimation of extent and determinants of intergenerational transmission of socioeconomic inequalities

Abstract This thesis investigates the extent and determinants of the intergenerational socioeconomic mobility. We first investigate the earnings transmission from fathers to sons in Germany, carefully addressing the question of biases in the estimation. However, this approach fails at taking account of all factors from the socioeconomic background of an individual affecting future success in life. We then consider sibling correlations as a broader indicator of all family influences, first in France, for education, profession and earnings. We also conduct a comparative study of the brother earnings correlation in France and Sweden to assess the impact on the estimation of the lack of information about permanent earnings and the use of predicted measures instead. Finally, we address the question of the mechanisms underlying the transmission of inequality. We thus explore the possibility to use the events of May 1968 in France as a natural experiment to identify and measure the causal link between parental and children's education.

Keywords: intergenerational mobility, inequality, earnings, education

Mobilité intergénérationnelle :
Une estimation internationale de l'ampleur et des déterminants de la transmission intergénérationnelle des inégalités socio-économiques

Résumé Cette thèse s'intéresse à l'ampleur et aux déterminants de la mobilité socio-économique intergénérationnelle. Nous nous intéressons d'abord à la transmission des revenus de pères en fils en Allemagne, en traitant soigneusement la question des biais d'estimation. Cependant, cette approche ne tient pas compte de tous les facteurs du milieu socio-économique d'un individu affectant sa réussite future. Nous considérons ensuite des corrélations au sein de fratries, comme indicateur plus large de toutes les influences familiales, d'abord en France, pour l'éducation, la profession et les revenus. Nous réalisons également une étude comparative des corrélations de revenus entre frères en France et en Suède, afin d'évaluer l'impact sur l'estimation de l'absence d'information sur les revenus permanents et de l'utilisation alternative de mesures prédites. Enfin, nous abordons la question des mécanismes sous-jacents à la transmission des inégalités. Nous explorons ainsi la possibilité d'utiliser les événements de mai 1968 en France comme expérience naturelle pour identifier et mesurer le lien causal entre éducation des parents et des enfants.

Mots clés: mobilité intergnénérationnelle, inégalités, revenus, éducation