SCIENCES PO

MASTER'S THESIS

Intergenerational Income Mobility in France: National and Territorial Estimates

Supervisor:
Prof. Pierre-Philippe COMBES

Author:
Gustave KENEDI

Co-Supervisor:
Dr. Clément DHERBÉCOURT
(France Stratégie)

May 22, 2017

Abstract

This paper provides new estimates of the extent of intergenerational income mobility in France, both at the national and subnational levels. Since we do not observe fathers' incomes, we follow a two-sample instrumental variable approach as used by Björklund and Jäntti (1997). Once we account for life-cycle bias in measures of son income, our estimates suggest the father-son intergenerational elasticity of earnings (IGE) in France is around 0.4-0.5, in line with the only available previous estimate. Moreover, we examine the spatial variations in social mobility within France. We find important variations in income persistence at the regional, departmental and urban area level. The main geographic pattern that emerges is that territories in the North, East and South-East exhibit relatively lower social mobility rates than territories located in the East and Center. Grouping departments into quintiles with respect to son income, we find that departments in the upper quintile display greater income persistence. Looking at the 10 largest urban areas in France, our results suggest the worst performing urban areas are Nice, Lille and Lyon, while the best performing are Grenoble, Strasbourg and Nantes. Future work should try to understand the underlying causes of spatial disparities in social mobility within France.

Acknowledgements

I am extremely grateful to Clément Dherbécourt for hosting me at France Stratégie and for assisting me throughout this project. This work would not have been possible without his help. I also want to thank Pierre-Yves Cusset, Pauline Grégoire-Marchand and Arthur Heim for stimulating conversations and thoughtful advice. Last, but not least, I want to extend my thanks to Pierre-Philippe Combes for accepting to supervise my dissertation and for his suggestions along the way.

This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissement d'Avenir" program (reference: ANR-10-EQPX-17 - Centre d'accès sécurisé aux données - CASD).

Contents

1 Introduction **3**

2 Measurement Issues **6**
2.1 Attenuation Bias 7
2.2 Age Bias 8
2.3 Lifecycle Bias 9

3 Empirical Strategy **9**
3.1 First Step 10
3.2 Second Step 11
3.3 Validity of Empirical Strategy 11

4 Data **11**
4.1 Sample Definitions 12
4.1.1 Sons 12
4.1.2 Synthetic Fathers 13
4.2 Variable Definitions and Descriptive Statistics 13
4.2.1 Variable Definitions 13
4.2.2 Descriptive Statistics 15

5 Results at the National Level **17**
5.1 Baseline Results 17
5.2 Extension: Age and Life-Cycle Bias 19
5.3 Robustness Checks 21
5.3.1 Sample Homogeneity 21
5.3.2 Outliers 22

6 Geographic Variation **24**
6.1 Regional Estimates 24
6.2 Departments 26
6.3 Largest Urban Areas 27

7 Conclusion and Suggestions for Future Research **29**
7.1 Future Research 30

8 Appendix **35**

1 Introduction

To what extent does socio-economic advantage persist from one generation to another? How enduring is intergenerational economic advantage? This question has been the source of a vast literature for a number of countries, predominantly the United States, first in sociology and since the 1970s in economics as well (Björklund and Jäntti, 2000). Sociologists have historically analyzed intergenerational mobility through the lens of social class and status, while economists have looked at measures relating to income.

The aim of this paper is to characterize the extent of social mobility in France from an economic perspective. Inspired by the remarkable recent work by Chetty et al. (2014a) for the United States and Dherbécourt (2015) for France, this paper goes further than providing national estimates and also investigates the spatial variations of intergenerational social mobility within France. This latter part of our contribution is still in its infancy and much work remains to be done to better understand the geography of social mobility in France.

There are two takeaways from empirical studies on the intergenerational transmission of income. First, there seems to be substantial variation in social mobility between countries. The intergenerational elasticity of income (IGE), the most commonly estimated measure of social mobility, varies from below 0.2 in some Scandinavian countries to close to 0.5 in the United States and the United Kingdom (Corak, 2013). These widespread differences can help identify policies that are linked with improved mobility and in this regard our study contributes to this endeavor. Second, measurement issues in both child and parent *permanent* earnings lead to substantial (mostly downward) biases. In particular, disentangling the income component that is transitory from that which is permanent is crucial and we devote considerable attention to these concerns.

Using data on children born in France on the first four days of October between 1970 and 1980, we estimate the intergenerational elasticity of income both at the national level and at various subnational levels. The major challenge in estimating these elasticities for France is that, currently, there is no available dataset which contains income data for both parents and their children. This adds another challenge to the already existing one regarding permanent income measurement. To overcome them we use a two-sample instrumental variable (TSIV) empirical strategy as proposed by Björklund and Jäntti (1997) in the context of IGE estimation, and previously employed by Lefranc and Trannoy (2005) in the French context. This two-stage method consists in predicting fathers' permanent earnings from a sample of "synthetic fathers" for whom income data is available, using occupation as our instrument. Our estimates suggest the father-son[1] intergenerational elasticity of income in France is most likely between

[1]In this paper, we focus our attention on father-son IGEs, leaving analyses regarding daughters and

0.4 and 0.5. In other words, a 10% increase in father permanent earnings is approximately associated with, on average, a 4-5% increase in sons' lifetime income. These estimates are slightly higher than those found by Trannoy and Lefranc (2005) which hovered around 0.40 for sons, though they are not significantly different. Perhaps one reason why our most robust IGEs are larger is that our dataset allows us to better control for life-cycle bias in the measurement of sons' incomes though it could also be due to methodological and data differences.

Moreover, we estimate IGEs at different subnational levels, an exercise which has not yet been done for France. Sons are assigned their location based on the town in which they were born regardless of where they moved later on. Because we are constrained by sample sizes, we look at both aggregated territories, regions, and densely populated departments (*Département*) urban areas (*Aire Urbaine*). This analysis suggests there are important spatial variations in social mobility within France, just as in the United States (Chetty et al., 2014a) and as already evidenced for France using social class (Dherbécourt, 2015; INSEE, 2016). In particular, intergenerational income persistence is lowest in regions located in the North, East and South East of France while those situated in the West and Center tend to be more mobile.

At more detailed geographic levels, we find some evidence that departments in the upper quintile in terms of population size or median son income are less mobile. One possible explanation may be that those areas also exhibit the highest levels of educational inequality fostering a strong intergenerational persistence of incomes.

Finally, we provide intergenerational elasticity estimates for the 10 largest urban areas in France. Nantes comes on top of the list with an IGE of around 0.23 while Nice appears as the urban area with the least social mobility with a persistence of 0.55. The three biggest French urban areas stand in the bottom third, with elasticities of 0.38, 0.40, and 0.49 for Paris, Lyon and Marseille - Aix-en-Provence respectively. It is important to note that in estimating these subnational elasticities we cannot adjust for life-cycle bias in son income and thus they should be interpreted with caution. In addition, they are based on small samples (between 250 and 900 observations, except for Paris) so standard errors are large and the differences are generally not statistically significant.

We hope these intriguing first results will stimulate further research on the spatial variation of social mobility in France and on understanding the underlying factors.

Related literature. The economic literature on intergenerational social mobility is extensive (see Solon (1999) and Black and Devereux (2011) for overviews). We restrict our attention here to the most relevant and salient results for the present study while methodological issues are detailed in the following section.

---

mothers for future research.

Despite much talk about meritocracy in France, very few studies have attempted to quantify intergenerational mobility with respect to income.[2] The only published paper on the subject, to the best of our knowledge, is Lefranc and Trannoy (2005)[3] and it is the one referenced by Corak (2013) and others (e.g. Grawe, 2006) for cross-country comparisons of social mobility. They use data from INSEE's five *Formation et Qualification Professionnelle* (FQP) surveys between 1964 and 1993 and predict fathers' permanent incomes following Björlund and Jäntti's (1997) application of two-sample instrumental variable. This method, as alluded to in the introduction, enables researchers to estimate IGEs even in the case when they cannot observe father income but have access to a sample of "synthetic fathers", that is a set of individuals for whom income data is available as well as common characteristics with actual fathers (such as education level, occupation, etc.). The instruments used by Lefranc and Trannoy (2005) are social class[4], following Erikson and Goldthorpe's (1991) classification, and education. They find that the intergenerational elasticity of earnings for France is between 0.34 and 0.44 for sons.[5] Their analysis is restricted to children aged between 30 to 40 years old when their income is observed and who report positive income and full-time full-year equivalent earnings above half the minimum wage. As such, France seems to stand at the higher end of developed countries with respect to the IGE, consistent with the assessment that France is the OECD country where education levels between generations are most closely linked (OECD, 2016).

These estimates at the national level, however, are likely to hide important variations at more detailed geographical levels. In a recent groundbreaking work, Chetty et al. (2014a) analyze the spatial variation in social mobility within the United States. Using administrative data on individual de-identified federal income tax records for approximately 10 million children, they detail how prospects of social mobility vary across 741 Commuting Zones (CZ). Their results suggest that there is substantial heterogeneity in social mobility between CZs, even more so than between countries. Exploiting these spatial variations, they conduct a correlation analysis to identify the characteristics of CZs most closely linked with better prospects for children from low-income households. Though we think conducting such an analysis would yield very insightful results, in this paper we only focus on estimating IGEs and leave the correlation analysis for future research.

Regarding France, there are two studies that look at the geographic variation of social mobility. The most comprehensive is Dherbécourt (2015). Using data from INSEE's

[2]That being said, the sociological literature is expansive. See Vallet (1999) and Peugny (2016).

[3]Arnaud Lefranc also has a working paper from 2011 analyzing the evolution of intergenerational income mobility during the mid-20th century in France.

[4]We will employ the words 'social class' and 'occupation' interchangeably.

[5]This range does not take into account standard errors which are not always negligible (between 0.026 and 0.061).

Enquête Emploi and using measures of social mobility across social class, he shows that across departments the prospects of children born to blue-collar and employee parents differ significantly. In particular, the proportion of these children who become managers and intermediate professions varies from 47% in Paris to 24.5% in Creuse (center of France). He also finds that chances of upward mobility are significantly increased for individuals who do not live in their birth region when they are adults though there are regional disparities. The second study is from INSEE (2016). Employing the same dataset as we use, but looking at occupation, the authors analyze differences in outcomes between children born in Ile-de-France, the region of Paris, and the rest of France. They find that regardless of parent occupation, it is easier for children born in Ile-de-France to become managers early in their careers than for those born in the rest of France. This can mainly be explained by higher educational attainments in Ile-de-France as well as more marked changes in the occupation structure in the region relative to other regions.

2 Measurement Issues

In this section we detail the main measurement issues that arise when estimating the intergenerational elasticity of earnings. On a sidenote, recent work by Chetty et al. (2014a) has shed light on the non-linearities of the IGE and sensitivity to observations with zero income, preferring to it measures based on the relative ranking of parents and children within their respective income distributions. Considering the limitations of our dataset and the fact that IGEs have been the standard in the literature, we remain with the IGE tradition. However, if better data becomes available for France this would be a fruitful avenue of research.

The intergenerational elasticity of earnings is one of the simplest and by far the most commonly estimated measure of social mobility in the economics literature. It relates the log of the child's[6] *permanent* income with the log of the father's *permanent* income in a linear framework.[7] Denote y_f and y_c respectively the log of father and child permanent earnings. The relationship between the two is described by:

$$y_c = \alpha + \beta y_f + \epsilon, \tag{1}$$

where α is the intercept, β is the elasticity of child income with respect to father income, our measure of interest, and ϵ is an error term assumed to be normally distributed. β

[6]The literature has predominantly focused on father-son IGEs and we do as well. We have obtained early results for daughters but more work is needed.

[7]Few papers explicitly justify the choice of using fathers' earnings rather than mothers'. Mazumder (2005) suggests using household earnings would provide a more accurate estimate. This is also the approach adopted by Chetty et al. (2014a, 2014b). We stay in the tradition of father-son IGE for comparison purposes.

represents the average percentage increase in a child's permanent earnings if his or her father's permanent income is increased by 1%. Another way of interpreting β is if two children's fathers have a 1% difference in their permanent earnings, they will, on average, have a β percent difference in their own permanent incomes. Importantly, the IGE is invariant to changes in mean income across generations, and thus measures *relative* mobility rather than *absolute* mobility (Torche, 2013).

As simple as this relationship may first appear, the literature has pointed out three important biases that have caused estimates to vary significantly: attenuation bias, age bias, and life-cycle bias (Black and Devereux, 2011). We explain each source of bias in turn.

2.1 Attenuation Bias

The first source of bias in estimating the intergenerational elasticity of income is attenuation bias. It arises when the measure for fathers' income captures *transitory* income rather than *permanent* income (Solon, 1992; Zimmerman, 1992).[8] This source of bias is crucial since in most datasets father income can only be measured in a single year. We can easily find the attenuation factor with a simple errors-in-variables setup. Assume $y_{f,t}$ is father's income in year t and y_f is his permanent income. The two are related in the following way:

$$y_{f,t} = y_f + \nu_{f,t}, \tag{2}$$

where $\nu_{f,t}$ is a transitory shock in period t and assumed to be uncorrelated with y_f (and y_c) and normally distributed. In this case, the probability limit of β from estimating equation (1) by OLS is

$$\text{plim}_{n\to\infty}\,\hat{\beta}_{OLS} = \frac{Cov(y_c, y_f)}{Var(y_f) + Var(\nu_{f,t})} = \beta\frac{Var(y_f)}{Var(y_f) + Var(\nu_{f,t})} < \beta. \tag{3}$$

Thus the estimated IGE is downward biased when there is measurement error in father permanent income.

Two solutions have been proposed to reduce this attenuation bias. The first is averaging fathers' incomes over a number of consecutive years. The second is using instrumental variable (IV) estimation.

Averaging fathers' incomes over T periods leads to an attenuation factor of

$$\frac{Var(y_f)}{Var(y_f) + \frac{Var(\nu_{f,t})}{T}} \tag{4}$$

[8] Note that the IGE is not biased by classical measurement error in the independent variable (children income) though as we will see it can be biased by more sophisticated forms of measurement issues for child earnings.

which tends to 0 as T increases. As better data becomes available and methodological insights improve, the recommended number of years over which father income should be averaged has varied. Solon (1992) averaged up to 5 years while more recent work by Mazumder (2005) has shown that the estimated IGE can be biased even with T greater than 10, depending on the extent of persistence of the transitory shock across time periods. However, it is yet unclear how applicable these results are to other countries than the United States (Black and Devereux, 2011, p.1491)

Since in practice most datasets only offer a single year of income observation for parents, researchers have looked at IV strategies. Solon (1992) and Zimmerman (1992) instrument fathers' permanent income by fathers' education or socioeconomic status. As Lefranc and Trannoy (2005) point out, the properties of the IV estimate will depend on the capacity of the instrument to capture inter-individual variance in permanent income. Moreover, the estimate will be *upward* biased if the instrument has a direct positive effect on child permanent income other than through father's permanent income. This condition is very likely to be true and thus the estimator provides an upper bound for the IGE (see Björklund and Jäntti (1997) for derivations).

As we will see in the following section, our empirical strategy is based on a variant of this IV method.

2.2 Age Bias

Age bias relates to the importance of the age at which child and father incomes are measured. We follow Black and Devereux's (2011) simple example to illustrate this bias. Suppose father and child incomes are measured at age a, $y_{f,a}$ and $y_{c,a}$ respectively. They can be defined as

$$y_{f,a} = \mu_a y_f + \eta_{f,a} \tag{5}$$

$$y_{c,a} = \lambda_a y_c + \eta_{c,a}, \tag{6}$$

where μ_a and λ_a represent the extent to which income measured at age a is representative of permanent income, and $\eta_{f,a}$ and $\eta_{c,a}$ are error terms. Equation (5) is a generalization of equation (2) where we had $\mu_a = 1$. This model enables incomes measured at certain ages to be better proxies for permanent earnings than at other ages. In this case, assuming the error terms are uncorrelated with each other and with permanent earnings, the probability limit of the estimated IGE is

$$\beta \left(\frac{\lambda_a \mu_a Var(y_f)}{\mu_a^2 Var(y_f) + Var(\eta_{f,a})} \right) = \beta \lambda_a \theta_a, \tag{7}$$

where $\theta_a = \left(\frac{\mu_a Var(y_f)}{\mu_a^2 Var(y_f) + Var(\eta_{f,a})} \right)$.

There are three main takeaways from this model. First, even in the case where $\lambda_a = \mu_a = 1$, the magnitude of the attenuation bias depends on the variance of fathers' transitory error term which itself depends on age. Evidence provided by Mazumder (2005) indicates that the variance of $\eta_{f,a}$ varies across the life-cycle and is at its lowest around age 40. This suggests that to minimize the size of our age bias we should attempt to obtain measures of fathers' income around their forties. Second, for small values of μ_a, θ_a can be greater than one. Thus, for measures of father income that are exceedingly far from permanent income due to the age at which it is measured, it is possible to have upward biased estimates of the IGE. Lastly, in this configuration, measurement error in child permanent income can lead to a bias if λ_a is different from one. As for the father, the magnitude of the bias will depend on the age at which income is observed.

2.3 Lifecycle Bias

The third important bias uncovered by the literature is life-cycle bias. This bias arises when λ_a and μ_a vary by age (and thus differ from 1). Using career earnings data in the United States, Haider and Solon (2006) estimate λ_a and θ_a over the life-cycle. Their results suggest that λ_a is small (close to 0.2) when men are around twenty years old and increases to close to 1 once they reach their thirties. It remains high up to their late forties and then decreases. This is consistent with the fact that high-income earners generally tend to have steeper wage profiles and as such differences in early career incomes between low and high permanent income individuals will tend to underestimate the true difference in permanent incomes. Black and Devereux (2011) stress that significant attenuation biases due to the non-representativeness of measured child income could arise if children are below 30 when their incomes are observed.

The estimates of θ_a follow a similar pattern. They are around 0.2 at age 20 and increase to 0.6-0.7 at age 30. Again, there is a plateau between the individuals' thirties and forties, after which, the estimate declines to 0.4-0.5 towards age 60. As with sons, these results suggest that it is best to measure father income between their thirties and forties, though even then the attenuation bias remains important (0.6-0.7).[9]

3 Empirical Strategy

Our dataset only contains information on fathers' occupations and we do not have any income data for them. However it does contain earnings data for individuals that are observed at the same time as fathers and for whom we also have their occupation. This

[9] See Grawe (2006) for additional details on life-cycle bias and its importance for international comparisons.

second group of individuals is our set of "synthetic fathers". In this case, estimating the IGE is still possible following a similar method to that employed by Björklund and Jäntti (1997) and subsequently in the context of France by Lefranc and Trannoy (2005). The only difference between our strategies is that our sample of "synthetic fathers" are individuals of our longitudinal dataset while they constitute groups of synthetic fathers from previous waves of their (cross-sectional) surveys. In practice this difference does not alter the estimation procedure.

The idea is simple and is composed of a two-stage procedure. Let $Z_{i,t}$ denote a set of socio-economic characteristics observed for both fathers ($i = f$) and synthetic fathers ($i = sf$) in period t. In our case, $Z_{i,t}$ is occupation. Fathers' and synthetic fathers' permanent income, respectively y_f and y_{sf}, can be expressed as

$$y_i = \gamma Z_{i,t} + \zeta_{i,t}, \tag{8}$$

where $\zeta_{i,t}$ is the transitory component of permanent wage not captured by current social class and assumed uncorrelated with $Z_{i,t}$. Substituting into equation (2) we obtain

$$y_{i,t} = \gamma Z_{i,t} + \zeta_{i,t} + \nu_{i,t}. \tag{9}$$

We detail below the precise steps of our estiation procedure.

3.1 First Step

We estimate γ from equation (9) using our set of synthetic fathers with OLS. To reduce the potential for life-cycle bias in the measure of father income, we restrict our sample of synthetic fathers to those aged between 30 and 45, as in Lefranc and Trannoy (2005), and include age and age-square in the regression.[10] We also exclude synthetic fathers reporting zero income or incomes in the top and bottom 1% of the income distribution. This is done to avoid certain individuals having a disproportionate effect on the coefficient for their occupation. The adjusted R^2 is smaller when the tails are included suggesting outliers add noise to the estimation. Fathers' permanent incomes are then predicted using $\hat{\gamma}$ obtained from the first stage.

Specifically, we separately regress synthetic fathers' log income on age, age-square and occupation for each birth year of our cohort (1970, 1971, up to 1980). After each regression, the predicted income by occupations is computed based on an age of 40. Each actual father is then assigned the corresponding income for his occupation in the year his child is born.

[10] We did run this first stage with interaction terms between occupation and both age and age^2 but this addition did not result in any significant change in either our estimates or model fit.

3.2 Second Step

Using the estimated permanent incomes of fathers, we estimate equation (1) by OLS. This yields

$$y_c = \alpha + \beta(\hat{\gamma} Z_{f,t}) + \psi_{f,t}, \tag{10}$$

where $\psi_{f,t}$ is the resulting error term.

All the IGE estimates we report in this paper are based on this two-step procedure. In the second stage we again exclude the top and bottom 1% of child earners. We come back to the importance of this choice in Section 5.3.2. In particular, we regress the log of son average income at ages 31 and 32 on the predicted permanent father income for each birth year of our cohort. When estimating the IGE for the entire cohort pooled together we include birth year fixed effects.

3.3 Validity of Empirical Strategy

One may be somewhat skeptical of the ability of this estimation method to yield reliable estimates. In particular, the use of occupation as our only instrument, since we do not have enough observations for education, could potentially introduce too little variation in father permanent incomes. Our answer is twofold. First, even if we could observe our fathers' education level, evidence from Lefranc and Trannoy (2005) suggests that our results would probably be more upward-biased than with only occupation since education potentially has a greater direct effect on child income than social class. Second, as occupation status can vary over time, for all our national estimates we provide IGEs for predicted permanent father income based on two reported occupations: at the birth of their child and in 1990. We think this provides a strong check on our results.

4 Data

We use data from the *Échantillon Démographique Permanent* (EDP), a longitudinal dataset containing information on individuals born during the first four days of October.[11] Unlike many longitudinal datasets, the EDP does not directly survey individuals at different stages of their lives. Rather it adds information on the individual from a number of sources. We take advantage of data contained in three such sources: 1) birth certificates, which contain information individuals' birth year and location, as well as parents' birth year and occupation that year; 2) the 1990 census, which contains information on father's occupation that year; and 3) the *Panel tous salariés*, a very large panel

[11] See Jugnot (2014) for an impressively detailed description of the dataset.

of French workers, from which we obtain income measures for our sons and synthetic fathers.

4.1 Sample Definitions

4.1.1 Sons

Our baseline cohort contains 20,446 men who (1) were born in metropolitan France (excluding Corsica) between 1970 and 1980, (2) whose father's social class is observed at birth and in 1990, and (3) whose average annual income between ages 31 and 32 is positive.[12] We exclude from our analysis sons of farmers and craftsmen, shopkeepers and business owners (*Artisans, Commerçants et Chefs d'Entreprise*) because income data for these occupations are only available from 2002 and 2009 respectively and, as such, it is not possible to predict incomes from our set of synthetic fathers for actual fathers with these occupations. Additionally we exclude sons of retirees or fathers without any professional activity because we cannot predict them a permanent salary. These 4 father occupations represent slightly less than 10% of our baseline cohort. Unless intergenerational earnings relations are substantially different for this excluded group, our results should hold for these individuals as well.

The reason why we wish to observe father occupation both at birth and in 1990 is that fathers may plausibly move between occupations so that their occupation at the birth of their child may not necessarily reflect their "lifetime" occupation, or at least the one which will enable us to mostly closely proxy their permanent earnings. As just discussed, this also provides us with a natural check on our results.

However, one may be concerned by the fact that to observe fathers' social class in 1990, the son must have still been living at home that year. This is a reasonable assumption for children born after 1973 (17 in 1990) but sons born between 1970 and 1972 may be different from the larger sample of sons for whom we observe father's social class at their birth. Comparing both the proportion of fathers in each occupations and the distribution of income at age 31-32, there are no striking differences between these two groups. This suggests our restriction does not substantially alter the composition of our sample. We come back to this issue in Section 5.3.1.

For the spatial analysis, we loosen requirement (2) and just keep sons whose father's social class is observed at birth. This increases our sample to 28,563 men. The reason why the criteria is loosened is to obtain a sufficiently large sample of children, without which our analysis would be seriously constrained.

[12] We also exclude individuals whose fathers were younger than 15 at their birth. This concerns only 13 individuals of our cohort and thus has no impact on our results.

4.1.2 Synthetic Fathers

Our sample of synthetic fathers contains all (1) male individuals, (2) aged between 30 and 45 in the different birth years of our cohort, (3) who report positive income in at least one of those birth years and (4) belong to an occupation corresponding to that of our fathers.[13]. The age range is chosen to minimize the potential for life-cycle bias in the estimation of permanent income of fathers, as discussed in Section 2.

Due to the increased workload induced by the 1990 census, the French statistical institute (INSEE) did not clean the wage dataset and thus we do not observe income in 1990. Therefore, to estimate the first stage equation for father occupations observed in 1990, we average synthetic fathers' income in 1989 and 1991 only keeping those who do not change social class over those 2 years.

4.2 Variable Definitions and Descriptive Statistics

4.2.1 Variable Definitions

Son income. Our measures of son income is average real taxable wage at age 31 and 32.[14] Taxable wage is defined here as wage net of all social security contributions (social security, retirement and unemployment), deductible and non-deductible CSG and CRDS.[15] This measure therefore only encompasses labor income and does not include any capital income that the individual may earn. Measuring income at the same age for all individuals of our cohort allows comparison between the different birth cohorts and pooling all birth years for the spatial analysis.

Synthetic father income. Income for synthetic fathers is defined in exactly the same way as for sons. As explained just above, since 1990 annual income is not observe, we use average income in 1989 and 1991 for synthetic fathers who do not change social class to estimate the first stage of fathers in 1990. Also note that before 2001, income is only available for individuals born in even years and so our group of synthetic fathers only includes, by definition, individuals that meet this criteria.

Father income. Fathers' incomes are not directly observed and are predicted following the First Step detailed in Section 3.1.

[13]Contrary to Lefranc and Trannoy (2005) we do not restrict our sample of synthetic fathers to individuals who report at least one child since that would decrease our sample size. However, synthetic fathers with and without children are relatively similar in terms of social class composition and earnings.

[14]2012 is the most recent year for which we have income data, hence our age bracket.

[15]CSG = Contribution Sociale Généralisée, CRDS = Contribution pour le Remboursement de la Dette Sociale

Synthetic father occupation. The occupation of synthetic fathers comes from the wage database and is reported by employers. We categorize them into 5 occupations: managers and highly qualified professions (3), intermediate professions (4), employees (5), high-skilled blue-collar workers (6)[16] and semi- and low-skilled blue-collar workers (7)[17]. If the synthetic father changes occupation in a given year, the one with the longest salary period is retained and if the latter is split even between the two jobs, the one with the highest wage is kept.

Except for the division of blue-collar workers into two separate occupations, the 1-digit classification is retained. The reason for this is 1) the 2-digit occupation classification differs too greatly between sources and there is no consistent way to harmonize them, 2) there is quite a bit of movement of workers between 2-digit occupations, and less at the 1-digit level, suggesting that the 2-digit occupation would be less able to capture the permanent component of income.

Note that the 1990 father may not necessarily correspond to the biological father as the Census question asks for the occupation of the "household father". Since we are interested in the environment of the child as he grows up, this has no particular significance for our analysis.[18]

Father occupation. Fathers' occupation comes from two sources. The one reported at the birth of the child comes from the birth certificate, and is self-reported by the father. The one reported in 1990 is taken from the 1990 Census. Again, the occupation is self-reported by the individual surveyed. The classification used is the same as for synthetic fathers.

Father birth year. Some father birth years are missing for individuals of our cohort. To avoid having to discard these fathers we fill in the gaps using data on fathers from the difference Censuses (1975, 1980 and 1990). As we just discussed, we cannot be 100% certain that we are correctly attributing father age since the Census questions do not specifically target the biological father but rather the head of household. This is only relevant for the descriptive statistic on father age at birth and has absolutely no incidence on our analysis.

[16]This corresponds to occupation code 61 ("Ouvriers qualifiés (n.d.a.)") for synthetic fathers between 1970-1980, occupation codes 62 ("Ouvriers qualifiés de type industriel"), 63 ("Ouvriers qualifiés de type artisanal") and 65 ("Ouvriers qualifiés de la manutention, du magaisnage et du transport") for synthetic fathers in 1989-1991, and 61 ("Ouvriers qualifiés de type industriel ou artisanal") and 65 ("Ouvriers qualifiés (manutention, magasinage et transport)") for fathers.

[17]This corresponds to all occupation codes starting with 6 other than the ones used for high-skilled blue-collar workers. For simplicity, we refer to this last category as low-skilled blue-collar.

[18]One could potentially use reported birth years in the birth certificates and the 1990 Census to identify biological fathers if one were particularly interested in genetic transmission.

4.2.2 Descriptive Statistics

Tables 1 and 2 present descriptive statistics on our sample of sons, fathers and synthetic fathers. Focusing first on our sample of sons, the mean of average income at age 31-32 is very similar across years, varying between 22,000 and 23,000. One important element is the sizable standard deviation for sons born in 1971, which is due to the fact that one individual born that year is a millionaire. We have also put the bottom and top 1% of the distribution to show that it is relatively close between each year.

Regarding our fathers, they are aged around 28 at the birth of their child. The proportion of fathers within each social class in the year of birth of child is very stable, with approximately 8.5% of managers and highly qualified professions, 17% of intermediate professions, 19% of employees, 31% of high-skilled blue-collar workers and 25% of low-skilled blue-collar workers. The occupation proportions in 1990 differ quite significantly, partly reflecting structural changes in the economy away from low-skilled blue-collar jobs. The percentage for this occupation decreases to around 19% and that for employees to 11%, while the proportion of managers and intermediate professions rises to 17% and 23% respectively. These changes are also suggestive that father occupation at birth may not be representative of the "main" occupation of our set of fathers, giving more force to our choice to examine father income predicted from both occupation at birth and in 1990.

One somewhat puzzling observation regards the slight mismatch in social class composition between our fathers and synthetic fathers. For each birth year, there is around a 10% difference in the proportion of employees between our sample of fathers at birth and synthetic fathers. Even if we restrict our attention to synthetic fathers with children and with similar ages as our actual fathers, the proportions do not match perfectly. These differences may be due to the varying reporting methods for both variables though this is difficult to verify. Father occupation, both at the birth of their child and in 1990, is self-reported, while synthetic father occupation is reported by the employer. Another explanation could be that our sample sizes are small and that the proportions are bound to not perfectly match. A slight difference between the social class of fathers and synthetic fathers remains in 1990, though for this year the divergence concerns managers and blue-collar workers.

Table 1: Descriptive Statistics on Sons and Their Fathers

	Sons				*Fathers*											
	Average Income at age 31-32					Social Class										
					Age at birth	At birth (%)					In 1990 (%)					N
	Mean	Std. Dev.	P1	P99		3	4	5	6	7	3	4	5	6	7	
Birth year																
1970	21,913	12,654	2,137	63,971	27.64	7.19	14.87	18.30	31.49	28.14	17.32	25.28	10.20	27.51	19.69	1,432
1971	22,425	30,757	1,464	61,542	27.70	7.08	15.80	19.10	31.49	26.53	17.98	23.53	10.67	29.54	18.28	1,696
1972	22,671	15,359	1,322	65,558	27.59	8.18	17.01	18.35	33.42	23.04	17.81	25.40	10.12	27.61	19.05	1,858
1973	22,106	15,375	1,776	62,584	27.89	7.78	16.36	17.87	31.37	26.62	16.63	22.95	11.56	28.73	20.14	1,852
1974	22,257	12,948	1,792	62,372	27.82	7.76	16.31	17.94	31.87	26.12	16.89	22.01	11.13	30.66	19.31	1,895
1975	22,264	11,182	2,261	58,402	27.68	7.84	16.57	18.48	32.14	24.97	17.41	24.36	9.91	30.18	18.14	1,786
1976	23,210	17,392	1,823	64,626	27.95	8.33	18.23	18.65	30.99	23.80	17.92	22.19	11.30	28.65	19.95	1,920
1977	22,579	13,719	1,573	63,297	28.28	8.12	17.08	19.91	30.97	23.92	16.98	24.77	10.30	30.06	17.89	1,873
1978	22,286	11,500	2,146	62,211	28.41	8.74	16.77	19.21	31.45	23.83	17.33	23.02	10.26	28.91	20.48	1,968
1979	22,519	12,608	2,343	57,821	28.68	9.19	18.09	18.92	29.67	24.12	18.04	23.15	11.04	29.86	17.90	2,056
1980	21,986	11,147	1,515	58,014	28.91	9.29	15.45	21.47	30.14	23.65	16.45	21.61	11.94	30.38	19.62	2,110
Baseline Cohort (1970-1980)	22,292	15,690	1,823	61,110	28.08	8.19	16.64	18.97	31.33	24.87	17.33	23.40	10.80	29.33	19.13	20,446

Notes: Son income is measured as the average income at age 31 and 32. Classification for father social class is: (3) - Managers and highly qualified professions; (4) - Intermediate professions; (5) - Employees; (6) - High-skilled blue-collar workers; (7) - Low-skilled blue-collar workers. See Sections 4.1 and 4.2 for sample and variable definitions.

Table 2: Descriptive Statistics on Synthetic Fathers

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1989-1991
Age	36.96	37.96	36.98	37.99	36.99	38.02	36.64	37.65	36.26	37.25	36.02	36.93
	(4.48)	(4.47)	(4.56)	(4.57)	(4.68)	(4.66)	(4.77)	(4.79)	(4.76)	(4.75)	(4.62)	(4,59)
Income	19,072	20,343	20,951	22,374	22,335	22,898	25,539	26,056	26,071	25,951	25,064	25,019
	(16,206)	(17,192)	(17,387)	(19,304)	(17,569)	(17,411)	(19,634)	(19,166)	(18,636)	(18,814)	(17,709)	(15,791)
Social Class (%)												
3	7.57	8.48	7.56	8.41	8.64	9.91	10.01	10.74	11.16	11.69	11.47	13.91
4	14.59	14.86	16.39	16.43	17.74	17.46	18.04	18.33	18.65	19.52	19.32	25.30
5	9.76	10.44	10.01	10.58	10.22	10.42	11.32	10.99	11.56	10.58	11.67	9.79
6	36.78	34.90	34.28	34.72	35.27	34.40	34.12	34.14	33.61	33.00	33.37	34.71
7	31.31	31.32	31.76	29.86	28.13	27.81	26.50	25.80	25.03	25.20	24.17	16.29
N	11,854	12,003	11,954	12,212	12,855	12,749	13,240	13,199	13,872	13,887	14,535	13,662

Notes : Classification for social class is: (3) - Managers and highly qualified professions; (4) - Intermediate professions; (5) - Employees; (6) - High-skilled blue-collar workers; (7) - Low-skilled blue-collar workers. See Sections 4.1 and 4.2 for sample and variable definitions.

5 Results at the National Level

5.1 Baseline Results

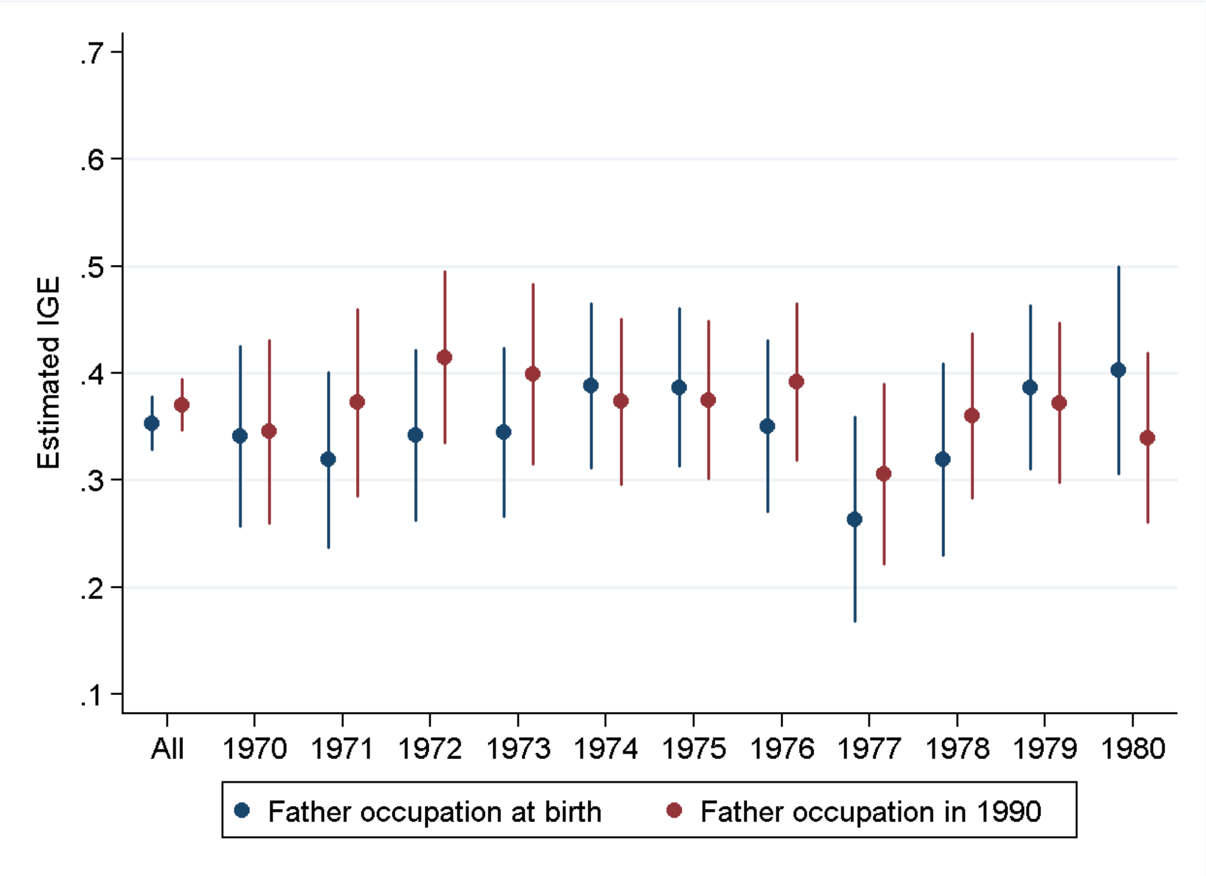
Following the two-step procedure laid out in Section 3, we first estimate the father-son intergenerational elasticity of income at the national level for France.

Before discussing the results, let us first briefly examine our first stage. Regression results from the first-stage can be found in Table 4 of the Appendix. Overall, the coefficients are very much in line with what we expect, i.e. positive for age and negative for age-square though not always statistically significant since the age range has always been constrained. Regarding occupations, managers and highly skilled professions earn most, followed by intermediate professions, employees, high-skilled blue-collar workers and low-skilled workers. One element worth highlighting is the relatively low fit of our first stage model. The adjusted R^2 are between 0.20 and 0.30, implying that most of the variance in synthetic father incomes is unexplained by our model. These fits compare with R^2 of around 0.50-0.54 in Lefranc and Trannoy's (2005) first stage regressions, though their regressions include education in addition to a more heterogeneous occupation classification (7 categories). This, in itself, is not a particularly problematic issue to the extent that occupation is able to capture a reasonably large proportion of individuals' permanent incomes (see discussion from Section 3.1).

To simplify the visualization of our estimated IGEs, we present coefficient plots and relegate regression tables to the Appendix. We indicate below each figure the corresponding regression table from the Appendix. For ease of comparability across figures we keep the axis scale constant.

Figure 1 presents the baseline results for our 1970-1980 cohort of sons for whom father occupation is reported both at birth and in 1990. In blue are the estimated IGEs

Figure 1: Estimated IGE for Entire Cohort and Individual Birth Years



Note: See Appendix Table 5 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval using robust standard errors

when permanent father incomes are predicted based on the occupation they reported at the birth of their child, while in red are the IGEs for when occupation is reported in 1990. The 95% confidence intervals, obtained from robust standard errors[19], are displayed to give a sense of the statistical proximity of our IGEs as well as to indicate the extent of uncertainty regarding our estimates.

The first two estimates, "All", are the IGEs when we pool each year of our cohort.[20] The estimated IGEs are 0.353 (0.013) and 0.370 (0.012) respectively when lifetime father income is imputed based on occupation at birth and in 1990. Reassuringly, they are not significantly different suggesting that they are at least robust to different measurement of father incomes. For each birth year, the IGE varies between 0.415 (0.041) and 0.264 (0.049) with all estimates, except these two extremes, between 0.3 and 0.4. 1977 appears to be a particular year, probably driven by sons outcomes since it is the smallest IGE for

[19]We would ideally have wanted to compute these standard error from a two-stage bootstrap as in Björklund and Jäntti (1997) but did not implement this method.

[20]Recall that for this estimation, we include birth year fixed effects.

both father income measures. As with the pooled cohorts, the difference between IGEs when father income is predicted based on reported occupation at birth and in 1990 are not statistically significant and tend to be very close to one another. No clear pattern is discernible throughout the studied period suggesting relative income mobility was more or less constant over the 1970s.

These baseline estimates are slightly smaller than those found by Lefranc and Trannoy (2005) though not significantly so. One reason may be that we measure our sons incomes at too early an age and thus our estimates could suffer from a downward life-cycle bias. Indeed, as we saw in Section 2.2 and 2.3, the IGE can be downward biased if the measure of child income at a certain age is not representative of his permanent income. Therefore in the following subsections we analyze in more depth how the IGE evolves when we vary both the age at which son income is measured as well as the number of years averaged to obtain the permanent son income measure.

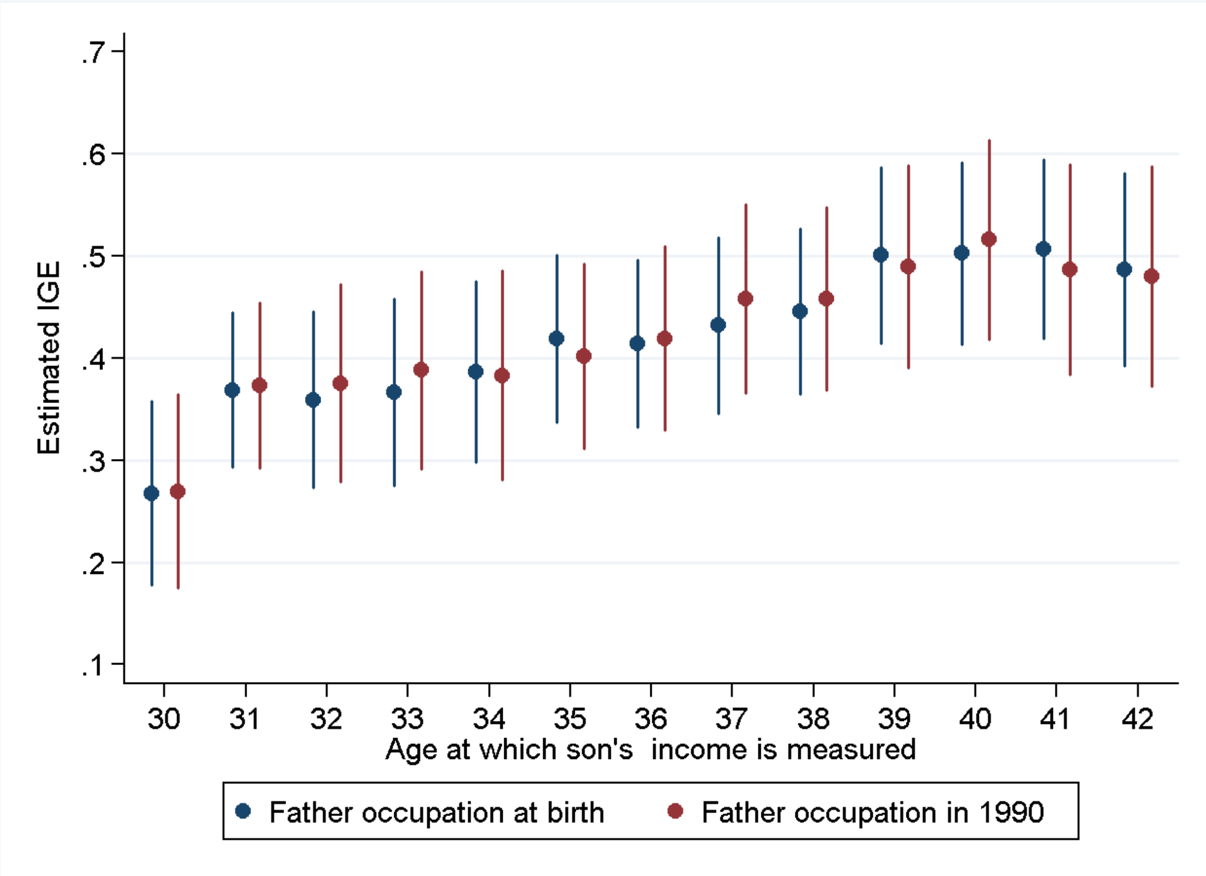
5.2 Extension: Age and Life-Cycle Bias

Our dataset does not allow us to directly verify whether our results are biased by life-cycle issues in fathers' incomes. However, given their incomes were predicted from (1) a restricted age range of synthetic fathers and (2) using age 40 as the prediction age, as recommended by the literature, we, *a priori*, believe this moderates the risk for such a bias. Moreover, the fact that IGEs based on reported father occupation at birth and in 1990 are very similar leads us to believe that life-cycle bias in father income is not a significant hindrance.

However, we do have reason to think that our income measure for sons may downward bias our results since we can only average over two years in the early thirties of our sons cohort to have a consistent income measure across all birth years. To check how our sensitive our estimates are to age and lifecycle bias in child income we conduct two extensions: (1) we vary the age at which son income is measured, and (2) we vary the number of years over which son income is averaged. For both, we only use sons born in 1970, the sample of sons tracked over time remains constant so that the IGEs are comparable, and we only include sons for whom we have income for those years (regardless of whether it is zero). For (2) the starting age is 30 such that averaging over 2 years means averaging income over ages 30 and 31. As with our baseline estimates, top and bottom 1% of earners are excluded. In total there are 918 sons for (1) and 982 for (2).

Figure 2 shows how the estimated IGE varies for measures of son income at different ages. The trend that emerges, and which reflects the age and life-cycle bias discussed in Sections 2.2 and 2.3, is that the IGE is a concave function of child age (conditional on a given father age). For our sample, the estimated IGE is below 0.3 when

Figure 2: Estimated IGE Varying the Age at which Son Income is Measured



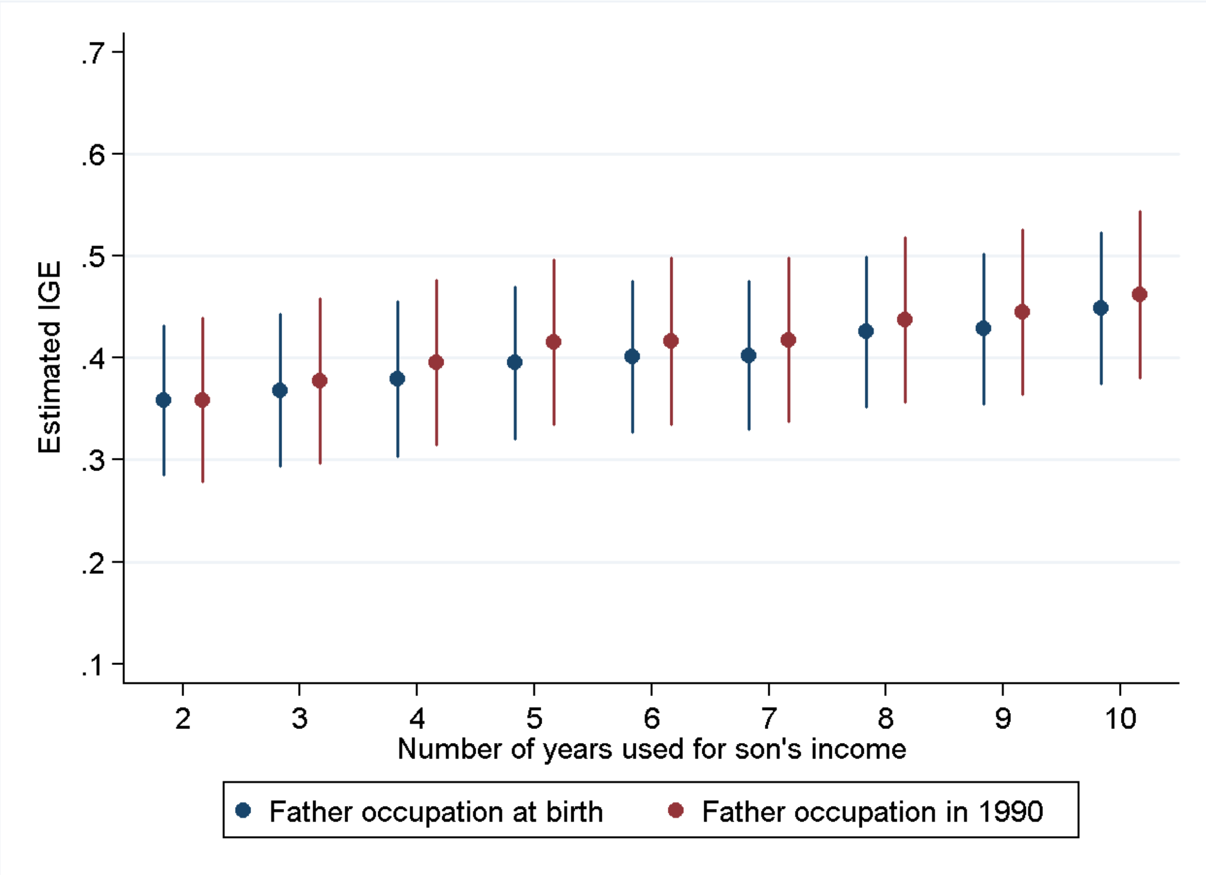
Note: See Appendix Table 6 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval, using robust standard errors. N=918.

individuals are 30 years old and increases up to 0.5 at 39 after which it remains at the same level and seems to start decreasing at 42 though we cannot, currently, evaluate what happens at later ages. Though there are no statistically significant differences between the estimates (excluding age 30 and probably due to the small sample size), these results are suggestive of a downward bias in our baseline results.

Figure 3 plots the estimated IGEs with respect to the number of years that are averaged to compute the measure of lifetime son earnings (starting from age 30). The same trend as in Figure 2 emerges: the estimated IGE increases as the number of years used to measure child is heightened. When averaging over only 2 years (30-31) the IGE is around 0.36 while it closer to 0.45 with 10 years of income (30-40). Just as in Figure 2, the differences are not statistically significant but strongly suggest that using a better proxy for permanent son earnings increases the IGE to between 0.4 and 0.5.

Overall, both extensions suggest our baseline results are likely to underestimate the extent of intergenerational income persistence due to life-cycle bias in son income.

Figure 3: Estimated IGE by Number of Years Averaged to Compute Son Income



Note: See Appendix Table 7 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval, using robust standard errors. N=982

In light of these refined estimations, we think the correct range of plausible IGEs for France lies between 0.4 and 0.5. Future work using higher quality data, if it becomes available, should confirm whether these results hold up to increased scrutiny.

5.3 Robustness Checks

In this subsection, we evaluate in turn the importance of two choices we made throughout the above analysis: sample restriction and outlier exclusion.

5.3.1 Sample Homogeneity

Our baseline cohort is comprised of children born between 1970 and 1980 for whom father occupation is available both at their birth and in 1990. This could potentially have led to increasing the homogeneity of our sample since it mildly favors men who left their home late in their youth and may be higher (or lower) income earners.[21] Although not discussed earlier, sample homogeneity was an early concern in the literature and led to downward biased IGE estimates (see Solon, 1992). Despite the fact that we could not discern any significant differences in terms of father occupations or incomes between our restricted cohort and the larger cohort of sons for whom father occupation is observed at their birth or in 1990, we estimate the IGE for these two larger son samples, pooling all years together for tractability.

Table 3 compares these IGEs with the ones previously found for our entire baseline cohort. The midpoint estimates are virtually identical with slightly smaller 95% confidence intervals for the larger samples. This suggests that our baseline cohort does not suffer from homogeneity bias and that our previous estimates are not affected by this potential concern.

Table 3: Sample Homogeneity Check

	Father occupation at birth	*Father occupation in 1990*
Sample of sons for whom father occupation is observed at both times	0.353 (0.328-0.378)	0.370 (0.346-0.394)
Sample of sons for whom father occupation is observed at their birth	0.352 (0.330-0.374)	-
Sample of sons for whom father occupation is observed in 1990	-	0.370 (0.347-0.393)

5.3.2 Outliers

Another choice we have made from the beginning is to exclude the top and bottom 1% of son earners. The reason is twofold. First, incomes at the top and bottom of the distribution possibly underestimate or overestimate lifetime earnings as people may currently be in short-term contracts or having a particularly good year. Second, from an interpretation point of view, we feel it is unreasonable to allow a very small group of individuals to drastically alter estimated IGEs. Indeed, most people would probably agree that it is not because a few individuals of either high- or low-income fathers have extremely high incomes that social mobility in France is dramatically greater or lower.

[21] Recall that to have the father's occupation in 1990, the son must have still lived with his parents.

Despite these justifications, and considering the literature has been relatively silent about the role of outliers[22], we check how much our estimates vary if we allow for outliers. Figure 4 shows estimated IGEs pooling the entire cohort together and for both measures of father income. Four specifications are plotted: with outliers in blue, excluding the top 1% in red, excluding the bottom 1% in green, and excluding top and bottom 1% in orange, our baseline specification which serves as a comparison benchmark.[23] The figure suggests there is some evidence for the importance of outliers at this aggregate level. Though the estimates are not statistically different the midpoint estimate drops from around 0.4 when outliers are included to around 0.35 when excluding the top tail. This is explained by the fact that our high income sons predominantly have high income fathers (i.e. managers and highly qualified professions). Interestingly, ex-

Figure 4: Estimated IGEs with and without Outliers

Note: See Appendix Table 8 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval, using robust standard errors.

[22]The issue of reported incomes of zero being problematic has been discussed by Chetty et al. (2014a). We have not (yet) tried imputing different values to individuals reporting zero incomes though this would provide another possible robustness check.

[23]Figures 9 and 10 in the Appendix present this exercise for each birth year for transparency.

cluding the bottom 1% of earners does not have any impact on the estimated IGE in our case.

6 Geographic Variation

In this section we present preliminary results on the spatial variation of social mobility in France. Sons are assigned at the location (town) in which they were born regardless of whether they moved later on. Our main constraint for this exercise is that we are working with a relatively limited number of observations. Therefore we pool together all sons born between 1970 and 1980 and for who we have father occupation at birth. We present results of social mobility at three geographic levels: regions, departments and urban areas.

Before presenting the results, we would like to stress that considering we are often working with relatively small samples, especially for departments and urban areas (between 250 and 1,000 observations), these results should be interpreted with caution.

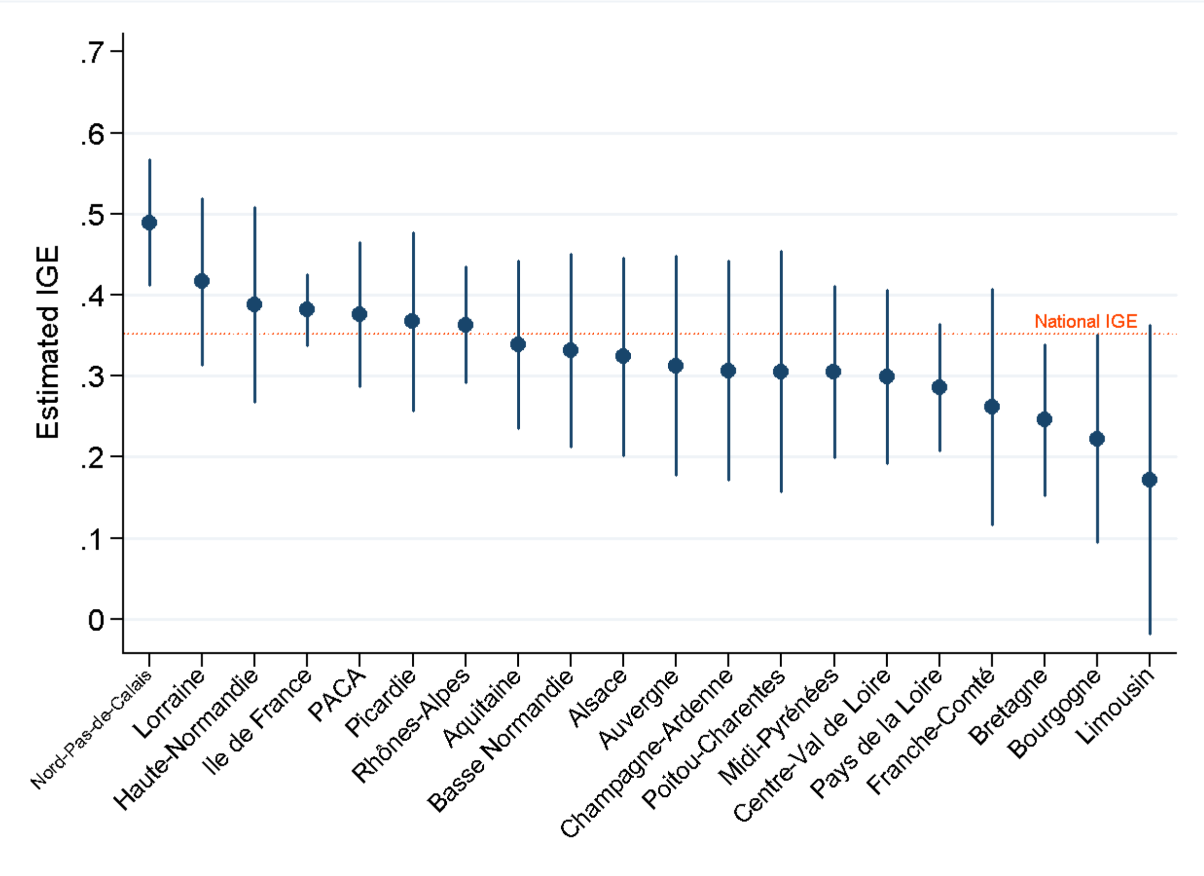
6.1 Regional Estimates

We estimate IGEs for 20 French regions, using the classification prior to the 2015 change that reduced the number of regions from 22 to 13. The 2 ommitted regions are Corsica and Languedoc-Roussillon as their sample size is below 250 observations. For each region we estimate equation (10) including birth year fixed effects, as we have done so far when pooling our cohort together, and excluding the top and bottom 1% of each region. The number of observation varies from 286 for Limousin (hence the large confidence intervals) to 5,668 for Ile-de-France, the Paris region. Except for Limousin, we have at least 600 observations per region. Descriptive statistics for sons and fathers for each region can be found in Appendix Table 10.

Figure 5 displays our results for the 20 largest French regions, in the same manner as we did for results at the national level. The estimated IGEs are ranked in decending order (i.e. from less mobile to most mobile). The dotted orange line represents the estimated IGE for the sample of sons used for these estimations (i.e. all sons for whom father occupation is reported at birth) at the national level and serves as a guide for comparison. Recall that our national IGE estimates suffer from a downward bias due to lifecycle issues in the measurement of son income such that regional IGEs should not be interpreted literally but rather should be employed to illustrate relative variations in social mobility across space.

Our results suggest that there are important variations in income persistence between regions, with IGEs ranging from close to 0.5 in Nord-Pas-de-Calais to around 0.2 in Bourgogne and Limousin. Interestingly, the largest French region, Ile-de-France

Figure 5: Estimated IGEs by Region



Note: See Appendix Table 9 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval, using robust standard errors.

exhibits relatively low rates of social mobility (0.382) and is situated in the bottom quintile of the distribution of regions. This is, in appearance, contrary to the findings of Dherbécourt (2015) and INSEE (2016) who overall found that children from Ile-de-France born to blue-collar and employee parents were more likely to become managers or intermediate professions. One reason may be that the IGE only captures the mean of the joint distribution of permanent incomes of fathers and sons. It could be that sons with poor fathers are enjoying higher wages but that around the middle of the distribution the persistence is very strong. Another reason may be that upward mobility in social class for this region is strongly driven by changes in the occupation proportions while incomes may not have reflected this structural change.

There are some spatial patterns that emerge. Regions located in the North (Nord-Pas-de-Calais, Picardie, Haute-Normandie and Ile-de-France), East (Lorraine, Alsace) and the South-East (PACA and Rhônes-Alpes) tend to have lower rates of social mobility while regions in the West (Bretagne and Pays de la Loire) and Center (Centre[24]

[24]There is a typo in the figure. It should be "Centre" rather than "Centre-Val de Loire" which is the

and Bourgogne) seem to enjoy greater social mobility than average.

In terms of socio-demographic characteristics[25], it is possible to uncover some tentative patterns. Populous regions tend to be towards the left hand side of the figure (e.g. Ile-de-France, PACA, Rhônes-Alpes and Nord-Pas-de-Calais) though there are a couple large regions towards the lower end (e.g. Bretagne and Pays de la Loire). In terms of regional GDP per capita, the richest regions (e.g. Ile-de-France Rhônes-Alpes and PACA) again tend to exhibit higher IGEs while poorer regions (e.g. Bourgogne, Franche-Comté and Limousin) have better outcomes though are a number of exceptions.

To get a more granular insight into subnational differences in social mobility rates, we now turn to the departmental level.

6.2 Departments

We conduct the exact same analysis for departments as we did for regions. Again, we restrict ourselves to departments with at least 250 observations once the top and bottom 1% of sons in terms of income are excluded. With this restriction, we are able to estimate IGEs for 41 departments. Even more so than for regions, results should be interpreted with caution as we are employing relatively small samples (between 300 and 600 observations in most cases).

Figure 6 displays our results for the 41 largest French departments. Just as with regions, it suggests that there are important (though generally not significant) variations in social mobility across departments in France. From a quick glance at the relative ranking of departments we do find the departments of regions at both extremes in similar positions. Figure 11 from the Appendix also shows similar spatial variations as for regions though the greater geographic precision unveils more subtle patterns.

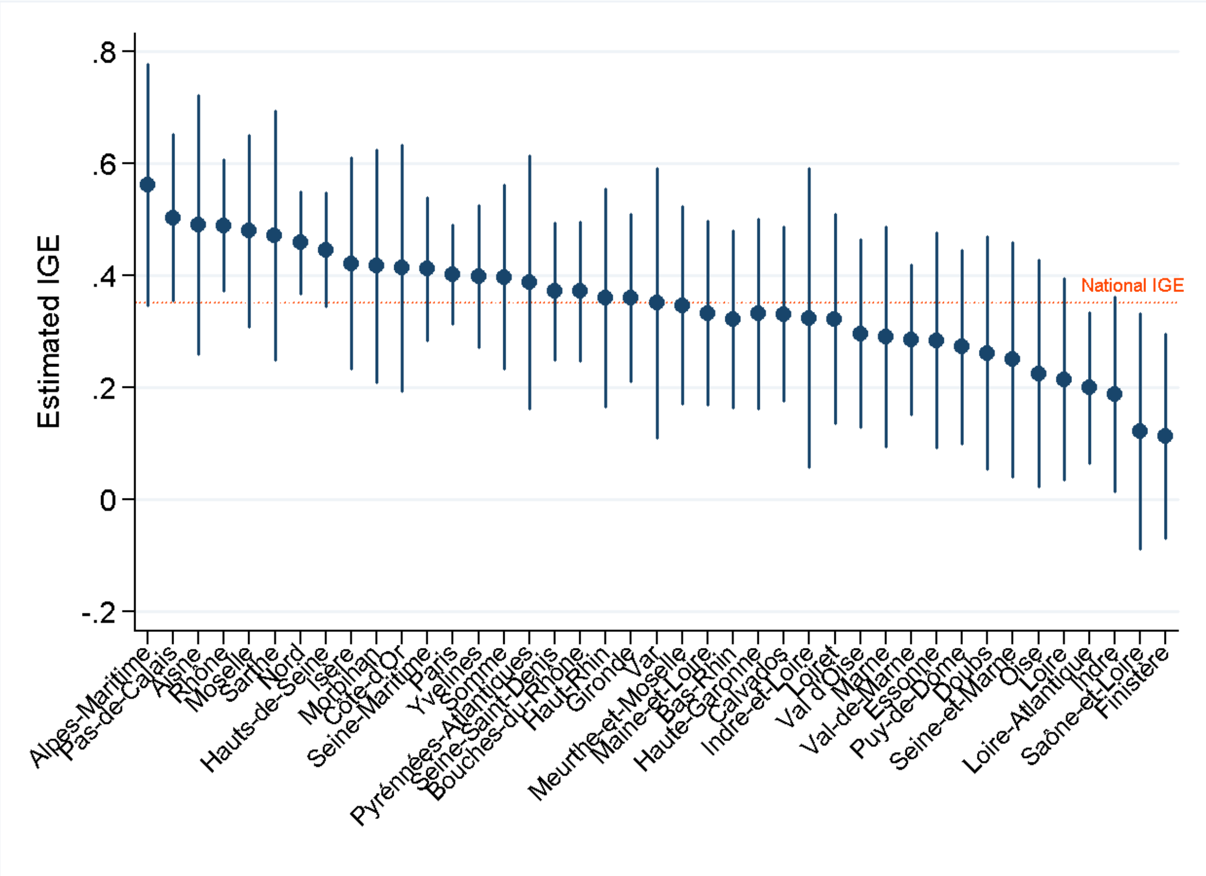
In an attempt to better understand the dynamics of spatial dispersion of social mobility, we group departments into quintiles with respect to 2 characteristics: (1) median son incomes, and (2) population size (using the number of sons in our sample). We thus re-estimate IGEs by quintile of departments for both measures. We provide descriptive statistics for each quintile in Appendix Table 12. Figure 7 displays the coefficient plot in the same fashion as we did for national estimates. q1 represents the lower quintile while q5 is the upper quintile. An interesting pattern emerges. Departments in the top quintile in terms of population and son income have lower rates of social mobility than other departments though the difference is only significant for quintiles by population size. For the latter in particular, the 4th quintile also emerges as relatively immobile.

It remains to be explained why more populous or richer departments have greater

new name of the region since 2015.

[25]Data from INSEE on the current state of regions is used for this simple analysis.

Figure 6: Estimated IGEs by Largest Departments



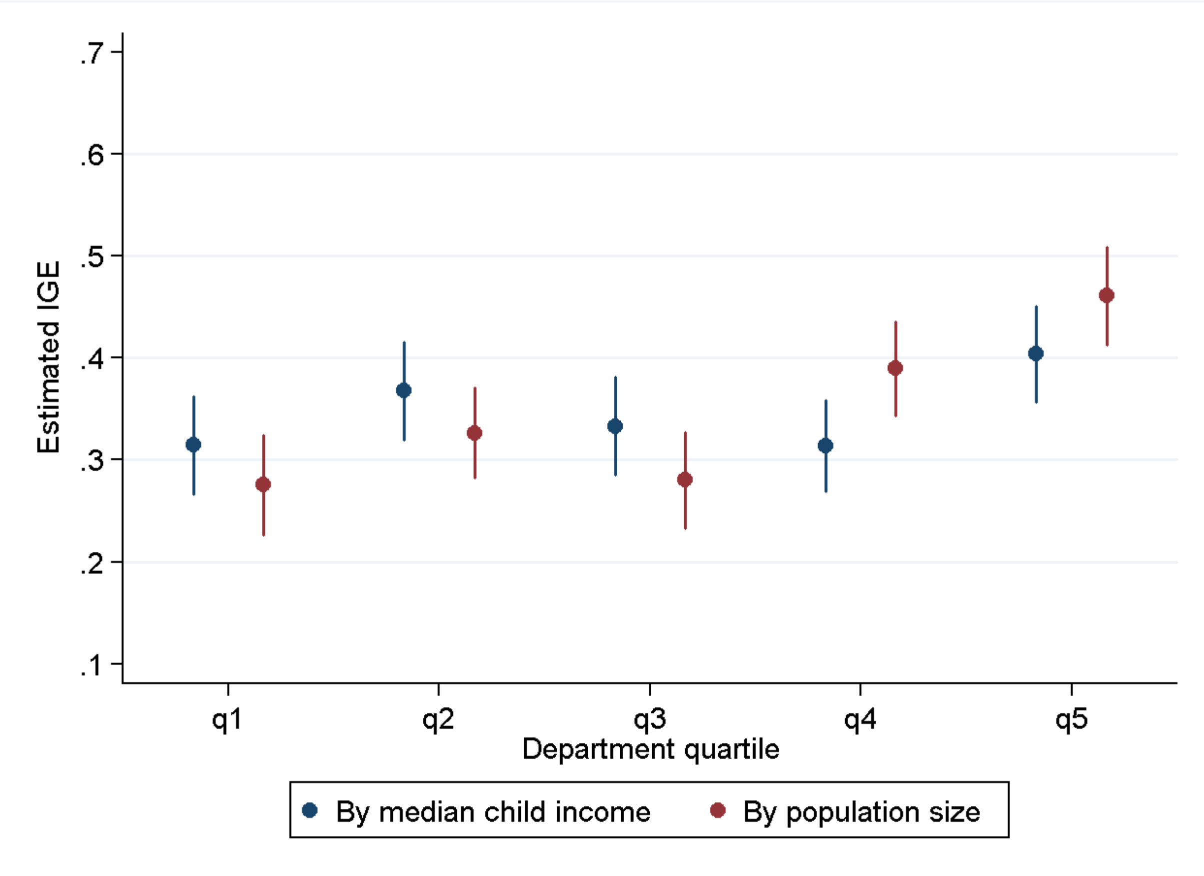
Note: We do not provide all regression outputs in the Appendix but they are available from the author. The circle represents the point-estimate while the lines show the 95% confidence interval.

intergenerational income persistence. One reason may be that the results are entirely driven by Paris, and we would have to re-estimate these quintiles excluding Paris to see whether the pattern remains. Another more general explanation, that we have not tested, is that children from less populous or wealthy departments will be more likely to move out to a city and thereby reduce the income difference between sons from poor and rich backgrounds. The unequal access to quality education may also play a role as overall richer areas will probably tend to have much larger inequalities in education quality.

6.3 Largest Urban Areas

Our last subnational exercise considers the 10 largest urban areas (in 2010). These are Paris, Lyon, Marseille - Aix-en-Provence, Toulouse, Lille, Bordeaux, Nice, Nantes, Strasbourg, and Grenoble. An urban area comprises the agglomeration as well as the neighboring towns of which at least 40% of residents work in the agglomeration or

Figure 7: Estimated IGEs by Quintile of Socio-Demographic Characteristics



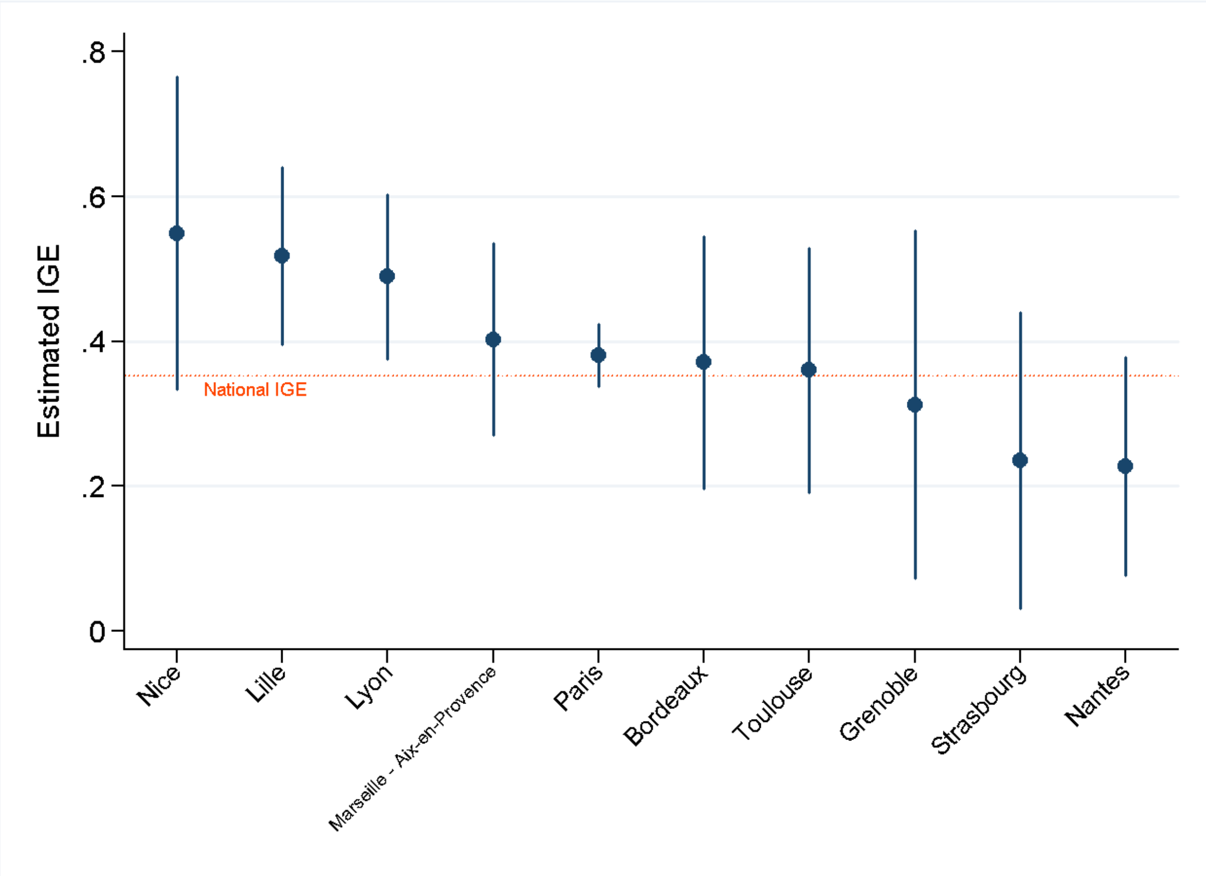
Note: See Appendix Table 11 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval, using robust standard errors.

towns that are "attracted" by it.[26] In general urban areas are much larger than the city itself and can encompass several departments, which is the case for the Paris urban area for example. Sample sizes are relatively constrained (except for Paris), with 6 urban areas having between 250 and 500 observations (see Appendix Table 14 for descriptive statistics). 11 individuals could not be matched to an urban area. We re-iterate one last time that IGEs should not be interpreted literally but in comparison with one another.

Figure 8 shows the results from these estimations, following the same format as our previous subnational figures. Nice, Lille and Lyon have IGEs around 0.50-0.55 while Strasbourg and Grenoble are between 0.20 and 0.22. In fact, the difference is statistically significant between Nantes and, Lille and Lyon. This is very much in line with the departmental results since the Nice urban area is composed primarily of towns in Alpes-Maritime, while the Nantes urban area almost exclusively encompasses towns in Loire-Atlantique. Paris does slightly less well than average as did its department

[26]This is INSEE's definition. See https://www.insee.fr/fr/metadonnees/definition/c2070

Figure 8: Estimated IGEs for 10 Largest Urban Areas



Note: See Appendix Table 13 for the detailed regression outputs. The circle represents the point-estimate while the lines show the 95% confidence interval, using robust standard errors.

and region.

The main takeaway from this urban analysis, very similarly to departments, is that large urban centers do not seem to promote social mobility, at least not measured by the IGE. The 3 largest French urban areas, Paris, Lyon and Marseille - Aix-en-Provence, are all in the first half of the worst performers, while the 3 smallest (of the 10 largest), Grenoble, Strasbourg and Nantes, all seem to enjoy high social mobility. Of course, these results are still preliminary and confidence intervals are particularly large considering our small sample sizes, so they should not be interpreted too literally but may be indicative of some overall pattern.

7 Conclusion and Suggestions for Future Research

This paper has used a novel longitudinal dataset of French residents born on the first 4 days of October to present new estimates of the intergenerational persistence of income in France. Our findings indicate that, once life-cycle bias in son's income is taken into account, the father-son IGE for France stands between 0.4 and 0.5. In international comparison, this compares with estimated IGEs of around 0.5 in the United States, the United Kingdom and Italy, and less than 0.2 in Finland, Norway and Denmark. As such, our estimates puts France is the group of developed nations with the least social mobility, at least as measured by the IGE.

We also document, inspired by the impressive work done for the United States by Chetty et al. (2014a), great spatial variations in social mobility within France. Because of the limited sample sizes with which we are working, we provide estimates at the regional, departmental and urban area levels when there are at least 250 observations. Regions in the North, East and South-East exhibit the lowest levels of social mobility while regions in the West and Center have the highest rates of intergenerational income mobility. At the departmental level, deviations from the national average are even greater. Reflecting the regional results, the 3 worst performing departments are located in the South-East and in the North, while the 3 best performing are situated in the West and Center.

To get a sense of potential underlying patterns, we group departments by quintiles of population size and median income. This very simple preliminary investigation suggests that social mobility in departments in the top quintile in terms of population and median income have greater intergenerational income persistence than departments in the lower quintiles, for which social mobility is relatively similar. This is consistent with our analysis at the urban area level where we find that the largest and wealthiest urban areas (Paris, Lyon and Marseille - Aix-en-Provence) tend to be less mobile than the smaller urban areas such as Nantes, Strasbourg and Rennes.

These suggestive subnational results call for greater attention and should be used in an attempt to determine local policies that are favorable to social mobility. There are numerous ways in which our research could be extended and we offer a number of suggestions in the following paragraphs.

7.1 Future Research

Analyze daughters

One big shortcoming of our paper is that we exclusively focus on father-son intergenerational income mobility. One straight-forward addition would be to replicate our analysis for daughters. We have actually already conducted an important part of the analysis for daughters but need more time to fully examine their implications. Indeed, our early findings suggest that social mobility is lower for daughters than for sons. Though the difference is not significant on year-by-year basis, these results differ from those obtained by Lefranc and Trannoy (2005) who found that daughters had better social mobility prospects.

Pool sons and daughters to obtain results for (almost) all departments

Another relatively straightforward extension would consist in pooling sons and daughters together to be able to estimate departmental IGEs for almost all departments. Though we would loose some of the particular differences between sons and daughters, we would gain on our ability to uncover spatial patterns and conduct a more rigorous correlation analysis to understand the underlying factors driving social mobility differences. An exercise similar to the one conducted by Chetty et al. (2014b) would no doubt prove highly insightful.

Exploit information on mothers and compute household income

Following the literature, we have only evaluated the influence of father incomes. Recent work by Mazumder (2005, p.236) suggests that family income "may be a better measure of parents' permanent economic status than fathers' earnings when only a few years of data are available." This is also the measure used by Lee and Solon (2009) and Chetty et al. (2014a, 2014b). Considering we have data on mother occupations at the birth of the child and in 1990 this could potentially be done. However, great care would have to be taken to predict incomes. Such an analysis could also enable us to examine the potentially differentiated influences of fathers and mothers between sons and daughters.

Estimate transition matrices

As we discussed, the IGE only portrays what occurs at the mean of the joint distribution of parent and child incomes. It says little about what takes place at different ends of this joint distribution. One alternative measure of social mobility that has been proposed is the transition matrix. The transition matrix divides both parental and child income distributions into quartiles or quintiles and provides the conditional probability of being in quantile i when being born to parents in quantile j. This statistics has not yet been, to the best of my knowledge, estimated for France while it has for the United States (Chetty et al., 2014b), Canada (Corak and Heisz, 1999), the United Kingdom (Blanden and Machin, 2008) and Denmark (Boserup, Kopczuk and Kreiner, 2013).

Considering the imperfections of our data, these conditional probabilities would have to be estimated with great care. We already know from O'Neill et al. (2007) that measurement error in parent or child income can substantially influence the predicted conditional probabilities. Nonetheless, we think this would be a worthwhile investment.

Exploit refined income data from FILOSOFI

The last extension we will suggest here is to exploit the extremely detailed income data available from the *Fichier Localisé Social et Fiscal (FILOSOFI)*, a newly added source to the *EDP*. For example, this new dataset gives us information on capital income and government transfers and could provide us with a more precise approximation of lifetime child income. These are just a few advantages of the *FILOSOFI* database that we could exploit.

Bibliography

Björklund, A. and Jäntti, M. (1997). Intergenerational Income Mobility in Sweden Compared to the United States. *American Economic Review*, 87(5), pp.1009-1018.

Björklund, A. and Jäntti, M. (2000). Intergenerational mobility of socio-economic status in comparative perspective. *Nordic Journal of Political Economy*, 26, pp.3-32.

Black, S. and Devereux, P. (2011). Recent Developments in Intergenerational Mobility. *Handbook of Labor Economics Vol. 4*, pp.1487-1541.

Blanden, J. and Machin, S. (2008). Up and Down the Generational Income Ladder in Britain: Past Changes and Future Prospects. *National Institute Economic Review*, 205(1), pp.101-116.

Boserup, S., Kopczuk, W. and Kreiner, C. (2014). Stability and persistence of intergenerational wealth formation: Evidence from Danish wealth records of three generations. *University of Copenhagen*.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014a). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4), pp.1553-1623.

Chetty, R., Hendren, N., Kline, P., Saez, E. and Turner, N. (2014b). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review*, 104(5), pp.141-147.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3), pp.79-102.

Corak, M. and Heisz, A. (1999). The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data. *Journal of Human Resources*, 34(4), pp.504-533.

Dherbécourt, C. (2015). *La géographie de l'ascenseur social français*. Document de travail. Paris: France Stratégie.

Erikson, R. and Goldthorpe, J. (1991). *The constant flux*. 1st ed. Oxford: Clarendon Press.

Grawe, N. (2006). Lifecycle bias in estimates of intergenerational earnings persistence. *Labour Economics*, 13(5), pp.551-570.

Haider, S. and Solon, G. (2006). Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, 96(4), pp.1308-1320.

INSEE (2016). *Gravir l'échelle sociale est plus aisé en Ile-de-France qu'en province*. INSEE Analyses N°50. Paris: INSEE.

Jugnot, S. (2014). *La constitution de l'échantillon démographique permanent de 1968 à 2012*. Document de Travail N°F1406. Paris: INSEE.

Lee, C. and Solon, G. (2009). Trends in Intergenerational Income Mobility. *Review of Economics and Statistics*, 91(4), pp.766-772.

Lefranc and Trannoy (2005). Intergenerational Earnings Mobility in France: Is France More Mobile than the US?. *Annales d'Économie et de Statistique*, (78), pp.57-77.

Mazumder, B. (2005). Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. *Review of Economics and Statistics*, 87(2), pp.235-255.

O'Neill, D., Sweetman, O. and Van de gaer, D. (2007). The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility. *The Journal of Economic Inequality*, 5(2), pp.159-178.

OECD (2016). *PISA France*. Note Pays. Paris: OECD.

Peugny, C. (2013). *Le destin au berceau*. 1st ed. Paris: La Republique des idées.

Solon, G. (1992). Intergenerational Income Mobility in the United States. *American Economic Review*, 82(3), pp.393-408.

Solon, G. (1999). Intergenerational Mobility in the Labor Market. *Handbook of Labor Economics Vol. 3*, pp.1761-1800.

Torche, F. (2013). *How do we characteristically measure and analyze intergenerational mobility?*. Stanford Center on Poverty and Inequality.

Vallet, L. (1999). Quarante années de mobilité sociale en France. *Revue française de sociologie*, 40(1), pp.5-64.

Zimmerman, D. (1992). Regression Toward Mediocrity in Economic Stature. *American Economic Review*, 82(3), pp.409-429.

8 Appendix

Table 4: First Stage Regression Results

Dependent Variable : log synthetic father income in year												
Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1989-1991
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	0.012	0.018	-0.001	0.030	0.0507**	0.064**	0.072***	0.061**	0.119***	0.093***	0.125***	0.045**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)	(0.020)	(0.0208)	(0.014)
Age-Squared	-0.000	-0.000	0.000	-0.000	-0.001*	-0.001**	-0.001**	-0.001**	-0.001***	-0.001***	-0.002***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Occupations												
4 - Intermediate	-0.365***	-0.409***	-0.382***	-0.396***	-0.402***	-0.379***	-0.385***	-0.412***	-0.374***	-0.395***	-0.366***	-0.401***
	(0.023)	(0.022)	(0.024)	(0.022)	(0.021)	(0.021)	(0.021)	(0.020)	(0.019)	(0.019)	(0.019)	(0.012)
5 - Employees	-0.840***	-0.913***	-0.882***	-0.875***	-0.881***	-0.811***	-0.777***	-0.779***	-0.763***	-0.768***	-0.727***	-0.720***
	(0.025)	(0.023)	(0.026)	(0.024)	(0.023)	(0.023)	(0.023)	(0.022)	(0.021)	(0.022)	(0.022)	(0.014)
6 - HS Blue Collar	-0.989***	-1.020***	-0.961***	-0.989***	-0.943***	-0.905***	-0.874***	-0.879***	-0.841***	-0.850***	-0.786***	-0.752***
	(0.021)	(0.019)	(0.022)	(0.020)	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)	(0.017)	(0.018)	(0.011)
7 - LS Blue Collar	-1.099***	-1.138***	-1.082***	-1.095***	-1.038***	-0.982***	-0.965***	-0.947***	-0.902***	-0.935***	-0.876***	-0.916***
	(0.021)	(0.020)	(0.022)	(0.020)	(0.020)	(0.019)	(0.020)	(0.019)	(0.018)	(0.018)	(0.019)	(0.013)
Constant	9.992***	10.13***	10.47***	9.985***	9.548***	9.293***	9.212***	9.400***	8.320***	8.774***	8.057***	9.494***
	(0.378)	(0.385)	(0.390)	(0.389)	(0.360)	(0.393)	(0.370)	(0.385)	(0.362)	(0.382)	(0.378)	(0.254)
Adj. R2	0.274	0.304	0.248	0.279	0.253	0.235	0.213	0.217	0.214	0.221	0.191	0.368
Obs	11,614	11,761	11,712	11,965	12,597	12,493	12,974	12,934	13,594	13,609	14,242	13,388

Notes : See Sections 3 and 4 for details. Robust standard errors in parenthesis.

Table 5: Regression Results for Figure 1

Dependent Variable : log average son income at age 31 and 32												
Year	All	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log father income predicted based	0.353***	0.341***	0.319***	0.342***	0.345***	0.388***	0.387***	0.351***	0.264***	0.319***	0.386***	0.403***
on reported occupation at birth	(0.013)	(0.043)	(0.042)	(0.041)	(0.040)	(0.039)	(0.0438)	(0.041)	(0.049)	(0.046)	(0.039)	(0.049)
Adj. R2	0.042	0.053	0.037	0.045	0.044	0.051	0.053	0.039	0.019	0.028	0.046	0.039
Log father income predicted based	0.370***	0.345***	0.372***	0.415***	0.399***	0.373***	0.375***	0.392***	0.306***	0.360***	0.372***	0.340***
on reported occupation in 1990	(0.012)	(0.044)	(0.044)	(0.041)	(0.043)	(0.040)	(0.038)	(0.037)	(0.043)	(0.040)	(0.038)	(0.040)
Adj. R2	0.049	0.049	0.044	0.060	0.052	0.047	0.055	0.058	0.031	0.045	0.051	0.038
Obs	20,036	1,402	1,662	1,821	1,815	1,857	1,750	1,880	1,836	1,928	2,015	2,066

Notes : See Sections 3 and 4 for details. Robust standard errors in parenthesis.

Table 6: Regression Results for Figure 2

Dependent Variable : log son income at age													
Age	30	31	32	33	34	35	36	37	38	39	40	41	42
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Log father income predicted based	0.268***	0.369***	0.359***	0.367***	0.387***	0.419***	0.414***	0.432***	0.445***	0.501***	0.503***	0.507***	0.487***
on reported occupation at birth	(0.046)	(0.039)	(0.044)	(0.047)	(0.045)	(0.042)	(0.042)	(0.044)	(0.041)	(0.044)	(0.046)	(0.045)	(0.048)
Adj. R2	0.043	0.097	0.075	0.076	0.073	0.101	0.102	0.112	0.123	0.120	0.117	0.120	0.105
Log father income predicted based	0.270***	0.373***	0.376***	0.388***	0.383***	0.402***	0.420***	0.458***	0.458***	0.489***	0.516***	0.486***	0.480***
on reported occupation in 1990	(0.048)	(0.041)	(0.050)	(0.049)	(0.052)	(0.046)	(0.046)	(0.047)	(0.046)	(0.051)	(0.050)	(0.053)	(0.055)
Adj. R2	0.038	0.085	0.071	0.074	0.063	0.081	0.092	0.109	0.113	0.100	0.107	0.097	0.089
Obs	918	917	918	918	917	917	917	917	917	917	917	917	918

Notes : See Sections 3, 4 and 5.2 for details. Robust standard errors in parenthesis.

Table 7: Regression Results for Figure 3

Dependent Variable : log son income averaged over X number of years

Number of Years	2	3	4	5	6	7	8	9	10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log father income predicted based	0.358***	0.368***	0.379***	0.395***	0.401***	0.402***	0.425***	0.428***	0.449***
on reported occupation at birth	(0.037)	(0.038)	(0.039)	(0.038)	(0.038)	(0.037)	(0.037)	(0.037)	(0.038)
Adj. R2	0.090	0.101	0.107	0.115	0.117	0.118	0.131	0.132	0.142
Log father income predicted based	0.359***	0.377***	0.395***	0.415***	0.416***	0.418***	0.437***	0.445***	0.462***
on reported occupation in 1990	(0.041)	(0.041)	(0.041)	(0.041)	(0.042)	(0.041)	(0.041)	(0.041)	(0.042)
Adj. R2	0.079	0.092	0.101	0.111	0.110	0.111	0.120	0.125	0.131
Obs	982	982	982	983	983	982	983	982	983

Notes : See Sections 3, 4 and 5.2 for details. Robust standard errors in parenthesis.

Table 8: Regression Results for Figure 4

Dependent Variable : log average son income at age 31 and 32

Outlier restriction	Including Outliers (1)	Excluding Top 1% (2)	Excluding Bottom 1% (3)	Excluding Top and Bottom 1% (4)
Log father income predicted based on reported occupation at birth	0.394***	0.348***	0.398***	0.353***
	(0.016)	(0.015)	(0.013)	(0.013)
Adj. R2	0.035	0.029	0.050	0.042
Log father income predicted based on reported occupation in 1990	0.417***	0.372***	0.414***	0.370***
	(0.015)	(0.015)	(0.013)	(0.012)
Adj. R2	0.042	0.035	0.057	0.049
Obs	20,446	20,241	20,241	20,036

Notes : See Sections 3, 4 and 5.3.2 for details. Robust standard errors in parenthesis.

Figure 9: Estimated IGE by Birth Year with Different Outlier Exclusions - Father Occupation Measured at Birth

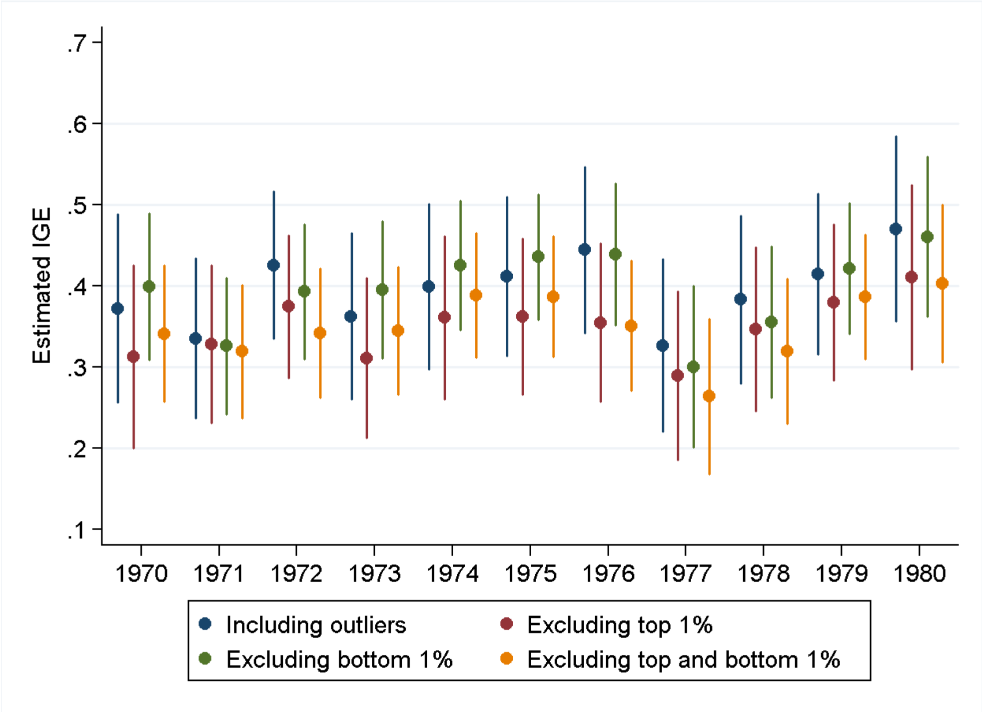


Figure 10: Estimated IGE by Birth Year with Different Outlier Exclusions - Father Occupation Measured in 1990

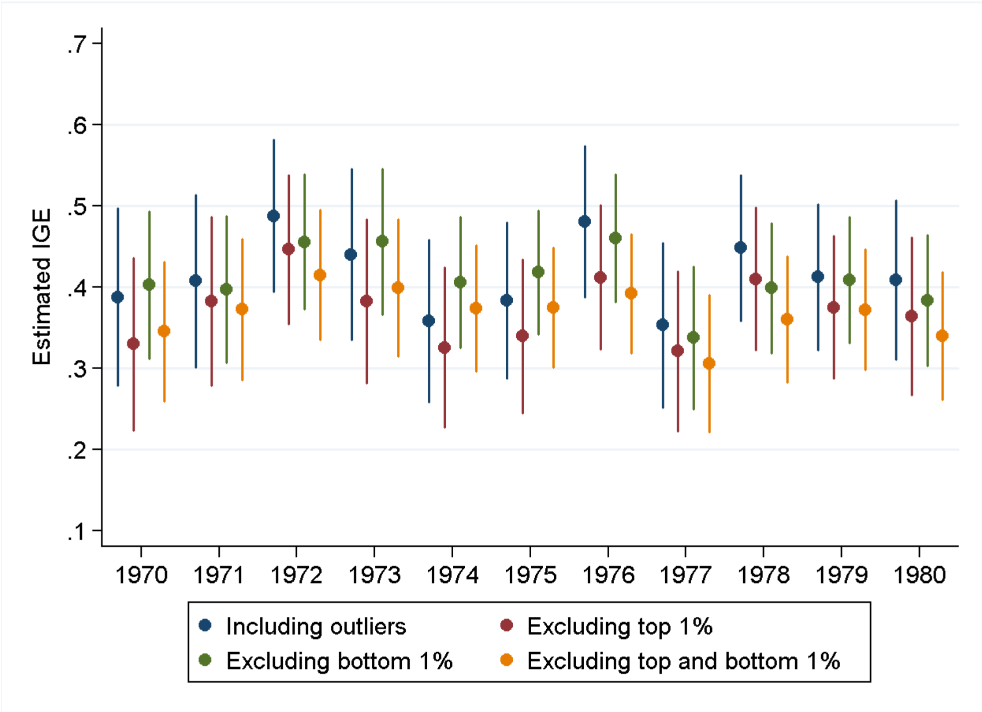


Table 9: Regression Results for Figure 5

Dependent Variable : log average son income at age 31 and 32

Region	IGE	Adj. R2	N	Region	IGE	Adj. R2	N
Ile-de-France	0.382*** (0.023)	0.049	5,668	Pays de la Loire	0.286*** (0.040)	0.031	1,706
Champagne-Ardenne	0.307*** (0.069)	0.026	784	Bretagne	0.246*** (0.048)	0.021	1,407
Picardie	0.367*** (0.056)	0.038	1,009	Poitou-Charentes	0.305*** (0.076)	0.011	744
Haute-Normandie	0.388*** (0.061)	0.034	1,020	Aquitaine	0.339*** (0.053)	0.032	1,239
Centre	0.299*** (0.055)	0.024	1,107	Midi-Pyrénées	0.305*** (0.054)	0.034	941
Basse-Normandie	0.332*** (0.061)	0.038	764	Limousin	0.172 (0.097)	0.036	286
Bourgogne	0.222*** (0.065)	0.018	833	Rhône-Alpes	0.363*** (0.037)	0.040	2,638
Nord-Pas-de-Calais	0.490*** (0.039)	0.055	2,616	Auvergne	0.313*** (0.069)	0.029	634
Lorraine	0.417*** (0.052)	0.045	1,314	Languedoc-Roussillon	0.277* (0.119)	0.027	225
Alsace	0.324*** (0.062)	0.040	785	PACA	0.376*** (0.045)	0.042	1,625
Franche-Comté	0.262*** (0.074)	0.014	617				

Notes : See Sections 3, 4 and 6.1 for details. Robust standard errors in parenthesis.

Table 10: Descriptive Statistics by Region

	Sons				Father		N
	Income at 31-32				Predicted income		
	Mean	Std. Dev.	P1	P99	Mean	Std. Dev.	
Region							
Ile-de-France	24,559	31,384	1,788	71,139	26,799	12,104	5,784
Champagne-Ardenne	20,956	12,359	1,825	59,286	21,920	9,151	800
Picardie	20,935	11,037	1,999	54,067	21,797	9,089	1,030
Haute-Normandie	21,342	12,814	1,523	61,107	22,100	8,835	1,041
Centre	21,617	13,638	1,998	60,127	23,140	9,793	1,131
Basse-Normandie	20,562	12,281	2,060	51,106	23,155	9,637	780
Bourgogne	21,227	10,902	2,211	54,788	23,526	9,835	851
Nord-Pas-de-Calais	20,777	10,919	1,251	56,394	23,197	9,805	2,669
Lorraine	22,572	14,256	1,788	68,547	22,948	9,314	1,342
Alsace	21,698	11,492	2,131	59,835	24,133	10,722	802
Franche-Comté	20,736	9,251	1,731	48,521	22,589	9,174	631
Pays de la Loire	20,927	9,580	1,787	54,729	23,450	10,047	1,741
Bretagne	22,432	11,852	2,086	62,978	24,069	10,536	1,436
Poitou-Charentes	21,134	11,869	904	59,206	22,752	9,483	759
Aquitaine	21,678	34,995	1,085	59,405	24,212	10,807	1,265
Midi-Pyrénées	21,242	10,915	1,485	59,138	24,525	10,851	960
Limousin	21,582	13,231	2,352	67,853	24,636	11,155	291
Rhône-Alpes	22,464	15,745	1,227	66,327	24,116	10,533	2,692
Auvergne	21,079	11,836	1,472	64,681	23,277	10,102	647
Languedoc-Roussillon	20,755	11,100	1,886	64,168	24,770	11,185	230
PACA	21,599	12,771	1,487	62,163	25,210	11,181	1,658

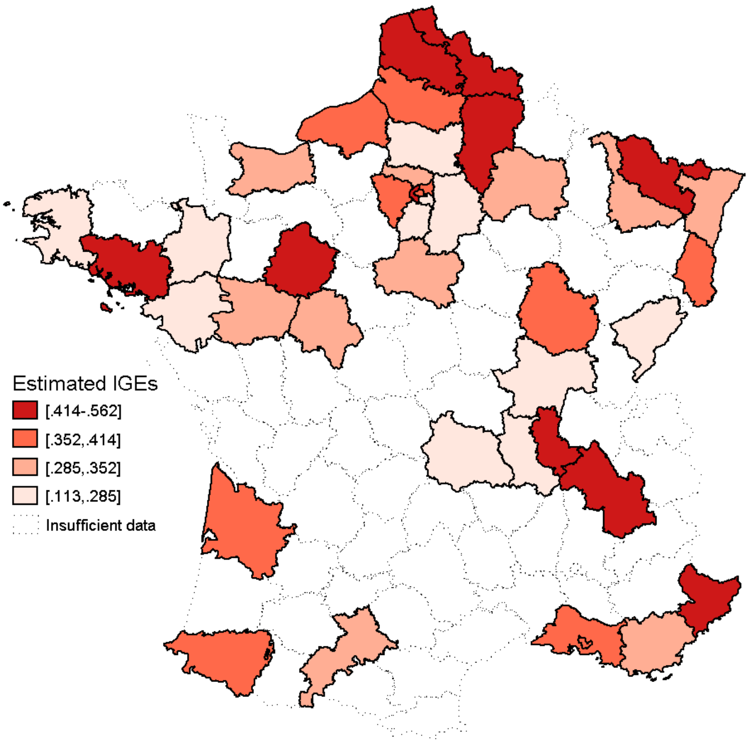
Table 11: Regression Results for Figure 7

Dependent Variable : log average son income at age 31 and 32

	Bottom Quintile	Second Quintile	Third Quintile	Fourth Quintile	Top Quintile
	(1)	(2)	(3)	(4)	(5)
Quintiles by Population Size	0.275***	0.326***	0.280***	0.389***	0.461***
	(0.025)	(0.023)	(0.024)	(0.023)	(0.024)
Adj. R2	0.023	0.034	0.023	0.047	0.068
N	5,630	5,993	5,662	5,797	4,891
Quintiles by Median Income	0.314***	0.368***	0.333***	0.313***	0.404***
	(0.025)	(0.025)	(0.025)	(0.023)	(0.024)
Adj. R2	0.029	0.038	0.035	0.029	0.056
Obs	5,826	5,725	5,310	6,349	4,737

Notes : See Sections 3, 4 and 6.2 for details. Robust standard errors in parenthesis.

Figure 11: Heat Map of IGEs for the Largest 41 Departments



Note: We do not report the 41 regression outputs for departments but they can be obtained from the author.

Table 12: Descriptive Statistics by Department Quintiles

	Sons		Father		
	Income at 31-32		Predicted income		N
	Mean	Std. Dev.	Mean	Std. Dev.	
Population Size					
Bottom Quintile	21,118	18,924	22,907	9,705	5,745
Second Quintile	21,367	12,682	23,385	10,055	6,117
Third Quintile	21,999	11,506	24,294	10,552	5,778
Fourth Quintile	22,646	14,967	24,666	10,797	5,917
Top Quintile	23,615	32,993	25,895	11,910	4,991
Median Income					
Bottom Quintile	20,683	11,626	23,202	10,014	5,945
Second Quintile	20,924	10,988	23,328	9,860	5,870
Third Quintile	21,792	19,828	23,819	10,375	5,420
Fourth Quintile	22,521	13,714	24,243	10,521	6,478
Top Quintile	25,068	34,218	26,719	12,208	4,835

Table 13: Regression Results for Figure 8

Dependent Variable : log average son income at age 31 and 32

Urban Area	Paris	Lyon	Marseille - Aix-en-Provence	Toulouse	Lille	Bordeaux	Nice	Nantes	Strasbourg	Grenoble
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IGE	0.381***	0.489***	0.402***	0.360***	0.518***	0.371***	0.549***	0.227***	0.235***	0.312*
	(0.022)	(0.058)	(0.068)	(0.086)	(0.063)	(0.089)	(0.110)	(0.077)	(0.104)	(0.122)
Adj. R2	0.049	0.070	0.062	0.058	0.084	0.034	0.064	0.007	0.034	0.008
N	5,712	919	679	359	723	454	280	409	299	312

Notes : See Sections 3, 4 and 6.3 for details. Robust standard errors in parenthesis.

Table 14: Descriptive Statistics by Urban Area

	Sons				*Father*		
	Income at 31-32				Predicted income		N
	Mean	Std. Dev.	P1	P99	Mean	Std. Dev.	
Urban Area							
Paris	24,531	31,270	1,816	71,018	26,730	12,090	5,830
Lyon	23,442	21,064	1,939	71,189	25,336	11,270	938
Marseille - Aix-en-Provence	21,946	12,032	1,746	63,386	25,557	11,693	693
Toulouse	21,781	11,029	1,137	58,751	26,842	12,032	366
Lille	22,656	13,493	771	58,353	25,476	11,257	739
Bordeaux	21,305	12,139	676	62,211	26,549	12,069	464
Nice	22,043	14,985	1,408	96,004	26,130	11,702	286
Nantes	21,795	10,578	1,729	55,446	25,308	11,253	419
Strasbourg	22,260	11,914	2,131	61,552	24,192	11,240	307
Grenoble	21,742	14,504	506	65,796	26,293	11,340	320