The Intergenerational (Im)mobility of Immigrants

Pascal Achard*

June 22, 2020

Abstract

This paper studies the influence of pre-migration social background on the long-term economic assimilation of immigrants. I use unique French survey data to trace family histories over three generations, before and after migration. While many immigrants experience an occupational downgrading at migration, their children benefit from the high socio-economic status their family had in the origin country. As a result, characteristics of immigrant grandparents are more predictive of their grandchildren's achievements than are characteristics of native grandparents. These findings are consistent with a model where immigrants cannot fully transfer their human capital between labour markets but transmit it across generations.

*Tilburg University. E-mail: p.p.m.achard@uvt.nl
I would like to thank my supervisors Andrea Ichino and Michèle Bélot for their support and their help. I also thank Jérôme Adda, Yann Algan, Inés Berniel, Leah Boustan, Agnès Charpin, Joseph Doyle, Johannes Fleck, Paola Giuliano, Pascal Jaupart, Fabian Lange, Guido Neidhofer, Sauro Mocetti, Eleonora Patacchini, Chiara Serra, Jan Stuhler, David Yang and participants at the Royal Economic Society 2017 Annual Conference, Associazione Italiana Economisti del Lavoro 2016 Annual Conference, the European Society for Population Economics 2018 Annual Conference, the 2017 OECD-CEPII Annual Conference on Migration, the European Association of Labour Economist 2018 Annual Conference, IZA Summer School 2018, the EUI Microeconometrics Working Group and the Louvain Macroeconomics Lunch Seminar for helpful comments. I am grateful to the "Réseau Quetelet" for having provided the data.

1 Introduction

Consider the following hypothetical example: a white-collar from a low-income country migrates to Europe. For many reasons (he does not write well in the language of the country, he cannot rely on his professional network, etc), he cannot find a similar job to the one he had before migration. He ends up working as a construction worker. In the destination country, he is indistinguishable from another immigrant construction worker who was an blue-collar worker before migration. Yet, they may be very different and their children may have different prospects in life. The objective of this paper is to investigate if this anecdotal story has empirical grounds.

To take such a long-term perspective, I use unique French survey data to trace family histories over three generations, in both the origin and destination countries, before and after migration. I use information on occupation and education of second generation immigrants (the children), their parents (who emigrated) and their grandparents (who stayed in the origin country). France has a long tradition of immigration with an estimated 40% of the French population having at least one immigrant grandparent[1]. This provides a large variety of origin countries and migration periods to study immigrants' long-term assimilation.

I find evidence that migration is a downward shock for many immigrants. I test whether the heterogeneity in economic conditions is U-shaped over time. This corresponds to a situation in which immigrants were heterogeneous before migrating, this heterogeneity is greatly reduced with the first generation and reappears with the second generation. There is clear evidence of such a pattern. Around half of the immigrants who had a managerial position suffer from occupational downgrading upon arrival. There is little catching up for the first generation. However, pre-migration status is key in explaining the educational outcomes of second generation immigrants. Having a father who prior to migrating had a managerial position is associated with (at least) 25% increase in the probability of finishing high school (for a given father's occupation in the destination country).

A strong (although lower) effect is associated with having educated or high socio-economic status (SES) grandparents. A 'grandparent effect' also exists for natives (Mare, 2011; Chan and Bolivier, 2013; Olivetti and Paserman, 2015), however it is different for immigrants. Since I restrict attention to immigrant grandparents who stayed in the origin country, they did not

[1] According to demographer Cris Beauchemin from the National Institute of Demography (INED) in charge of the survey used in this paper. See interview in https://www.la-croix.com/France/Immigration/Quelle-part-limmigration-demographie-francaise-2018-02-21-1200915395.

suffer from downgrading. Information about their status gives a more accurate signal than information on parents. Therefore, grandparents' characteristics should be more important in predicting immigrant outcomes than it is in predicting natives'.

It is difficult to compare the importance of family background since the distribution of education levels and occupations differ substantially between countries and time periods (Xie and Killewald, 2013; Long and Ferrie, 2013). I use machine learning techniques, i.e. random forests (Breiman, 2001), which allow to compute the relative importance of each variable in terms of its predictive ability. I show that grandparents' education and occupation as well as parents' education are more informative (with respect to parents' occupation) for immigrants than natives.

I also show that immigrants' intergenerational mobility appears lower when measured with origin (rather than destination) country characteristics. Depending on the classification used, mobility is estimated to be between 12 and 17% lower. All these results do not vary substantially with origin country, year of and age at migration.

The data allows to go beyond descriptive evidence and unveil mechanisms. It contains information on the transition between labor markets (first job and professional experience in France) and parental investment in children (schooling decision, help with homework, etc). I formalize the intuitive example above in a model, built around the idea that immigrants cannot entirely transfer their human capital (HC) between countries but can transmit it across generations. In the hypothetical example, it would mean that the white collar who became a construction worker still raises his children as a white collar would.

First generation immigrants have difficulties accumulating back human capital after the initial loss of migration. Their returns to labor market experience in the destination country are low. However, they do not suffer from the same downward shock when it comes to transmitting HC to their children. While on the labor market part of their parents' HC was lost, at home the children can fully benefit from it. Returns to investment in children are high making it easier for second generation immigrants to accumulate HC.

The rest of the paper is organised as follows: section 2 reviews the literature and section 3 presents the data. Section 4 shows the results on long term persistence of family background. Section 5 outlines a model and tests empirically some of its predictions. Section 6 concludes.

2 Literature review

The first strand of literature this paper contributes to is the one on the intergenerational mobility of immigrants. Most studies (Borjas, 1993; Aydemir et al., 2009) relate outcomes of first and second generation immigrants in the destination country. These papers control for origin country but do not include family specific information on pre-migration status. Origin country dummies may not capture well individual situations if immigrants from the same country are heterogeneous.

Recent work (Ward, 2020; Abramitzky et al., 2019) studies the long-term social mobility of immigrants in the U.S. Ward (2020) follows immigrants over three generations and finds strong persistence of grandparents' characteristics on grandchildren, stronger than predicted by a standard grandfather-grandson elasticity. Abramitzky et al. (2019) focuses on immigrants' mobility over two generations in two different periods of U.S. history; the Age of Mass migration and the current period. They find a higher rate of upward mobility at the bottom of the earnings distribution for immigrants than for natives. I depart from these two studies in that I take the situation of the family in the origin country as the relevant starting point to assess the mobility of second generation immigrants (the children).

In sociology, Ichou (2014) studied the effects of parents' relative (in the origin country) education level on the decision to migrate and on their children's educational achievements. He finds support for positive selection into migration and a positive effect of parents' relative level of education on the success of their children[2]. Catron (2020) studies the influence of pre-migration background on the economic success of second generation immigrants in the U.S. The author links ship manifest records (which include occupation before migration) to U.S. census. He finds a large effect of pre-migration status on the second generation (although lower than on first generation's outcomes).

The second strand of literature this paper contributes to is the one on immigrants' economic assimilation (Friedberg, 2000; Eckstein and Weiss, 2004; Cohen-Goldner and Eckstein, 2008).

[2]Part of my paper deals with questions not tackled in Ichou (2014), namely the loss of human capital at migration and the difficulty for first generation immigrants to catch up. Including them makes the link between the situations of different generations clearer. I also go beyond descriptive evidence and show through which channels family pre-migration characteristics operate. On the topics that overlap between the two papers, I go a step beyond in several key dimensions. I look at the effects of grandparents' characteristics on second generation immigrants and not just at parents'. This highlights long-term persistence. I benchmark the mobility of immigrants with that of natives and show that grandparents matter more for immigrants than for natives. I recalculate mobility statistics using origin country information and show how different it is from only looking at family characteristics in France.

Caponi (2011) develops a dynamic model where the first generation is ready to suffer a loss for the benefit of the second. The author matches survey data from the US and Mexico to estimate a structural model of immigrants' earnings. In this paper, I use data which directly follows the same families before and after migration. This allows to better apprehend the (dis)continuities brought by migration at the individual level.

Finally, this paper contributes to the general discussion on long term social mobility (Chan and Bolivier, 2013; Mare, 2011; Braun and Stuhler, 2018). This literature typically finds less mobility across multiple generations than was first thought by Becker and Tomes (1986). It is an open question whether it is driven by the presence of older generations (typically grandparents) or by the transmission of latent unobservable characteristics (Clark, 2014). This paper exploits a setting in which presence of older generations does not play a role as I focus on families where grandparents did not migrate. This paper provides evidence supporting the latter explanation.

3 Description of the data

3.1 One dataset, two samples

The dataset used in this paper is "Trajectoires et Origines" (TeO), collected by the National Institute of Demographic Studies (INED) and the National Statistical Agency (INSEE) in 2008/2009. It is a cross sectional survey[3] (18,864 persons interviewed in total) based on representative samples of immigrants and natives. TeO has several advantages over alternative data sources such as labor force surveys and censuses. First, TeO is designed with the specific purpose of studying first and second generation immigrants; it asks questions targeted for this population. Second, TeO specifically samples 2nd generation immigrants. People do not appear as immigrants in civil registries if they are born in France. Second generation immigrants had to be sampled from registries of first generation immigrants. Last but not least, TeO contains information on the socio-economic status in the country of origin.

Three different populations are surveyed in TeO: (i) first generation immigrants, (ii) second generation immigrants and (iii) natives. The questions in the survey refer to different moments in

[3]Abramitzky et al. (2014) raises two concerns about the use of cross-section data: they do not allow (i) to observe return migration and (ii) to account for the quality decrease of immigrants' cohorts (in the U.S. case). In the context of this paper, temporary migrants are not a population of interest, the relevant one is made of those who settle and have a family in France. When necessary, I include year of arrival fixed effects to show robustness to different immigration cohorts.

the lives of the people being interviewed: the personal and family history, the current situation and the situation of the children[4]. I do not have to merge different data sources to gather information on different generations (Caponi, 2011; Abramitzky et al., 2014, 2019).

The sample that surveys first generation, from now on referred to as *sample first*, contains mostly information on the situation prior to migration and on the parents' situation in France. The sample that surveys second generation, from now on referred to as *sample second*, contains mostly information on the situation of the children and on parental investments in their children[5].

I use the following notation to refer to immigrants' situation: $S_{j,t}$ describes the status of immigrants j in period t. Status refers to the information on education, occupation or household resources described below. The generations who grew up in the origin country are $j = P$ for parents (and grandparents), the generation which grew up in France is $j = C$ for children. There are three periods: $t = 0$ for information pre-migration, $t = 1$ for the time of migration and $t = 2$ for present days.

The data is composed of immigrants who arrived in different years and at different ages. I perform robustness checks to confirm that the results are not driven by differences in migration periods. In everything that follows, second generation immigrants are people whose both parents are immigrants. They are referred to as the children, the population referred to as parents and grandparents are theirs.

3.2 Information on the situation in the origin country

In *sample first* I observe parents' and grandparents' level of education and occupation. I use these measures to classify $S_{P,0}$[6]. I use a categorical variable taking 6 values for occupation and one taking four values for education[7]. For ease of presentation, occupation categories are sometimes transformed into a binary variable where the two categories with highest earnings

[4]Table A1 shows how different sections of the questionnaire are combined to generate two samples of second generation immigrants. By relying on two samples instead of one, I use all the information available in the survey. I can also check that my results are consistent across samples. I do not match the two surveys and always use them separately.

[5]Figures A1 and A2 show a digram (respectively for *sample first* and *sample second*) of who is being interviewed and how they relate to the population of interest. Using *sample first* allows to stack three generations. This is why it is the main sample used in this paper, contrary to Ichou (2014) which relies uniquely on *sample second*.

[6]Strictly speaking information on grandparents is measured at $t = 2$. However, grandparents' education is more than likely to have been finished before their children migrated (since parents migrated at adulthood). Information about occupation is elicited though a general question about grandparents' occupation and not one that specifically refers to period $t = 2$

[7]Education levels are categorized as below primary school, primary school, secondary education and higher education. Following the system used in France, occupations are grouped into six categories: self-employed agricultural, self-employed non-agricultural, high managerial, supervisory occupations, lower services and lower technical.

”high managerial” and ”supervisory occupation”, are merged into a ”high” occupation.

In *sample second* The only information available is parents' education.

Sample restrictions (detailed with figures A1 and A2) ensure that parents and grandparents acquired their human capital in the origin country.

3.3 Information on the situation in the destination country

Information on parents' situation In *sample first*, parents' situation is measured in three ways: the occupation in $t = 1$, just following migration, the occupation in $t = 2$ at the time the survey was taken and the total monetary resources available to the household in $t = 2$. In *sample second*, parents' situation can be assessed by their occupation in $t = 2$.

Outcome variables for children In *sample first*, children's education level can be inferred through a question asked to their parents. They have to report if their children passed the baccalauréat, France's end of high school exam. $S_{C,2}$ is a binary variable taking value 1 if children passed the baccalauréat.

In *sample second*, more outcome variables can be observed, in particular dropping out of school, obtaining the baccalauréat or a higher university degree

Sample restrictions (detailed with figures A1 and A2) ensures that children acquired their human capital in France.

Potential mechanisms What I call mechanisms in this paper are parents' investments in developing the human capital of their children. In *sample second*, children are asked whether in their youth, (i) they attended a different school than most children in their neighbourhood (investment labelled *schooling strategy*), (ii) had additional private lessons, (iii) received help from their parents with their homework and (iv) had a room of their own to study.

To minimize recollection errors (Braun and Stuhler, 2018), I use answers to questions concerning the preceding or following generation (parents about grandparents, children about parents, etc). Measures of family background should capture permanent status to avoid biases due to measurement error (Black and Devereux, 2011). Parents' educational outcomes rarely change at adulthood. Parents' occupation in France is also stable (as shown in the next section). Grandparents' occupations are elicited through a general question, not one referring to a particular point in their lives. Parents' occupation just before and after migration only capture a snapshot of their situation. These are, however, crucial periods for the research question addressed in this paper. All results are robust to using different measures of family status, which alleviates concerns about measurement error issues.

3.4 Descriptive Statistics

Table 1 provides basic descriptive statistics on first generation immigrants in *sample first*. I report the gender of the person being interviewed together with the distribution of their educational level and that of their father (the grandfather from the perspective of second generation immigrants), their birth year, the year in which they migrated and their occupation before migrating. The first two columns show this information for all immigrants (raw numbers and percentages), while subsequent columns break down immigrants by their region of origin (Europe, North Africa, Africa and Asia)[8].

There are two main takeaways from this table: first, there is variation in the pre-migration characteristics. The education level of first-generation immigrants roughly splits in three groups of equal size: below primary school, primary school and above primary. Grandfather occupation is concentrated in the lower technical category (45%), yet all categories have non trivial shares of people working in them. The same holds for grandparents' education and parents' pre-migration occupation. Among immigrants having a job before migration, 22% were "high managerial" or "supervisory occupations"[9].

The second takeaway is that there is a critical number of observations for each grouping of region of the World and education level/occupation. Immigrants from North Africa have the lowest education levels, yet 23% of the people reporting an occupation had a "high" one. African immigrants have the highest education levels, yet 21% of those with a job before migration reported working as a "lower technical".

[8] Table 1 provides statistics for immigrants with adult children. Table A2 reports the same information for all immigrants (with and without adult children). In table A3, a similar table is produced for the parents of immigrants followed in *sample second* (with the information available for that sample). The two samples are very different and correspond to different waves of migration (as can be seen from the distributions of arrival years).

[9] The number of first generation immigrants not reporting an occupation before migration is very large because the sample reports characteristics of males and females. In what follows, I restrict the sample to fathers when I focus on pre-migration occupation.

4 Resurgence of pre-migration characteristics

4.1 Compression of socio-economic heterogeneity following migration

Migration is a persistent negative shock for a large share of immigrants. To show that the shock is negative, I document how the heterogeneity in socio-economic status (SES) that existed in the origin country was reduced at the time of migration. To show that the shock is persistent, I document the mobility in SES between the year of arrival in France and the time of the interview. I use observations from *sample first* since they contain information on SES at the three relevant periods, before migration, at arrival and now.

Table 2 represents transition matrices from $t = 0$ to $t = 1$ and table 3 from $t = 1$ to $t = 2$. In these tables, $S_{P,0}$ is measured with father's pre-migration occupation[10], $S_{P,1}$ and $S_{P,2}$ with occupation at arrival and at the time of the interview. The upper part of the table reports conditional probabilities, the lower part reports raw numbers. In the last two columns, occupations are merged into a binary variable "high" and "low".

The main takeaway of table 2 is the compression of pre-existing heterogeneity. The majority of immigrants start their professional life in France working in a lower technical occupation. This holds particularly for immigrants having a "low" occupation before migrating, whether or not it is lower technical. It also holds for a large share of immigrants with a skilled occupation, "high managerial" and "supervisory occupation". 47% of high managerial, arguably the highest category, could not find a similar job upon migration. In the same fashion, 55% of immigrants with a supervisory occupation end up in a lower job.

The main takeaway of table 3 is that occupational mobility is low for immigrants. Although some diagonal elements are relatively low (lower services to lower services is 31%, supervisory occupation to supervisory occupation is 60%), mobility is limited to blocks within the matrix. There is for instance transition between high managerial and supervisory occ., between lower services and lower technical. Mobility numbers are much lower when grouped into binary categories, they are never below 70% and 5 out of 6 are above 80%.

In appendix, I address potential concerns arising from immigrants arriving from different countries and at different periods[11]. Transition between $t = 0$ and $t = 1$ is specific to immigrants

[10] To avoid issues related to female labour participation, I restrict the sample to males.

[11] To see if the results vary by origin countries or year of arrival, I run two regressions (reported in table A4) where the dependent variable is a dummy for low occupation occupation in $t = 1$ (Panel A) and in $t = 2$ (Panel B) and the main regressor another dummy variable respectively for $S_{P,0}$ and $S_{P,1}$. I run the regressions without

but transition from $t = 1$ to $t = 2$ can be compared between immigrants and natives. I show in appendix that mobility is lower for immigrants[12].

4.2 Reappearance of heterogeneity with the second generation

I test if pre-migration background has an effect on children's outcomes above and beyond father's occupation in the destination country. Table 4 reports the results of the following regression which resembles an AR(2) model

$$(S_{C,2})_i = \alpha + \beta_1 (S_{P,0})_i + \beta_2 (S_{P,2})_i + \gamma' X_i + \varepsilon_i \tag{1}$$

Subscript i refers to children and X is a vector of controls. In the baseline specification, the controls are age and gender. Standard errors are clustered at the family level[13]. The second column of panels A and B reports the coefficients of $S_{P,2}$, a categorical variable (taking 6 values) for father's current occupation. The first column reports the coefficient of $S_{P,0}$. In panel A, it is measured with father's pre-migration occupation; in panel B, with grandfather's occupation[14]. I report the F test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$ below their respective column. Note that coefficients in both columns are taken from the same regression.

Pre-migration characteristics reappear strongly with the second generation. All the coefficients in the first column are positive, suggesting that only previous advantage (and not disadvantage) reappears. Individual coefficients are significant at the top, meaning the pre-existing advantage reappears for the highest occupation categories (for parents and grandparents). The results are particularly strong for self-employed non-agricultural and high managerial positions (for both fathers and grandfathers) held in the origin country. The coefficients should be read as follows, for a given father's occupation, having a father who worked as a high managerial before migration is associated with an additional 29 p.p. probability of obtaining the baccalauréat for

controls in the first column and add successively year of birth fixed effects, year of arrival fixed effects, age at arrival and origin country. If the coefficient of interest (dummy for $S_{P,0}$ and $S_{P,1}$) varies a lot with the inclusion of extra regressors, this is indicative that the result is driven by the conditions at the time of arrival. The coefficients are completely insensitive to adding year of birth, year of and age at arrival. They change with the inclusion of origin country fixed effects, although the magnitude of the change is small, in particular for the transition between $t = 1$ and $t = 2$.

[12] In table A5, I regress the dummy variable for having a low occupation at $t = 2$ on a dummy for having a low first job interacted with immigration status. This tells how much occupation mobility differs between natives and immigrants. In the second column, I include year of birth fixed effects to account for starting professional life in France at different ages. A F-test of equality between immigrants and natives strongly rejects the null (value of the test above 40).

[13] Since I observe potentially several children per parent, as in Catron (2020)

[14] Grandfather from the father's side

an unconditional average of 58% of success. This means a 50% increase in education achievements. This number is still high, 22 p.p. when one considers having a grandfather working as a high managerial.

These numbers are also large when compared with the second column, i.e. the advantages associated with father's current occupation. For each category, coefficients of current and pre-migration status are within the same order of magnitude (when $S_{P,0}$ is measured by father's pre-migration occupation).

In tables A6, I replicate the same analysis but include sampling weights. In table A7, I include country of origin fixed effects and, in table A8 I use information from *sample second*. The main elements of the picture remain unchanged. In tables A9 and A10, I replicate the analysis but change the measure used for $S_{P,2}$. It is respectively first occupation in France and total resources available to the household at the time of the interview. The findings are robust to these different choices of $S_{P,0}$, $S_{P,1}$ and $S_{P,2}$. They constitute the main result of this paper.

4.3 Benchmarking the results with natives

The literature on social mobility which combines information over three generations has shown that a 'grandparent effect' is not specific to immigrants (Mare, 2011; Chan and Bolivier, 2013; Olivetti and Paserman, 2015). To be consistent with the story of this paper, it should be that the influence of grandparents' (or parents' characteristics before migration) are more important for immigrants than for natives. Father occupation is a poor indicator for immigrants and other (origin country) family characteristics should have more predictive power. It is difficult to make such comparisons when using linear regressions. The distribution of education levels and occupations differs between countries with different levels of development. For instance, having a grandfather who finished primary school means something different (an should be allowed to have a different effect) for a native and an immigrant.

To circumvent this problem, I use machine learning techniques (random forests, Breiman (2001)) to calculate (and compare) the relative importance of parents' and grandparents' characteristics in predicting children's achievements for immigrants and natives. I do not use the values taken by a variable but the importance of that variable in predicting the outcome. I do not compare if, for a given father's occupation in France, having a high managerial grandfather has a stronger effect for an immigrant than for a native. Instead, I see if the variable "grandfather occupation" has more predictive power for immigrants than for natives.

Random forests are classification techniques which construct a multitude of decision trees to predict an outcome. These techniques have the particularity of not only splitting the sample into a training and test datasets but also randomizing the variables which are used for splitting each tree. Random forests are very popular in machine learning because of the easiness of their implementation and the robustness of their results. Since the algorithms change the variables used to train the forest, it can also be used to assess the relative importance of each variable[15]. I use this relative importance measure to assess if family background is more important for immigrants than for natives.

Results on variable importance are usually normalized by the most important variable and ranges between zero and one. In this paper, I normalize variable importance by the variable measuring father's occupation in France. This calculates how much grandparents' characteristics and parents' education are informative (with respect to father's occupation) for immigrants and natives. This choice of normalization follows from father's occupation being the only measure taken at the same place and time by natives and immigrants. Results are reported in table 5.

In the upper panel, I include the following variables in the forests' algorithm: age, gender, grandfathers' (from both sides) occupations, parents' and grandparents' (of the parent being interviewed) education and father's occupation in France. In the lower panel, I drop the information on parents' education. I use one model for natives, reported in the third column and two models for immigrants, one with and one without a categorical variable for origin country. In this table, the value 1 means as informative as father's occupation. What is important is to compare the numbers between immigrants and natives. Larger values for immigrants means that the same information is more important in predicting immigrant children's outcomes.

For all background measures and in both panels, family characteristics are more informative for immigrants than for natives. In the algorithm where origin country is an explanatory variable, grandfathers' occupations have a predictive power equivalent to 0.98 and 0.92 of father's occupation when it is only 0.89 and 0.81 for natives. The differences are even larger when using information on immigrant's education, 0.98 and 0.9 for father's and mother's education versus 0.73 and 0.79 for natives.

[15]For an introduction, see Biau and Scornet (2016). The criteria used to calculate variable importance is the change in mean decrease impurity. There is, to my knowledge, no established way to calculate standard errors for these predictions

In table A11, I reproduce the analysis with data from sample second. Fewer family characteristics can be included (only parents' education) but their influence can be tested on several different outcomes (dropping-out, having the baccalauréat and obtaining a higher education degree). For all the outcomes and for the two models (with and without country of origin), family background is more informative for immigrants than natives. In *sample second*, parents' education is not only more informative for immigrants than natives, it also has a higher predictive power for immigrants than father's occupation (i.e. values greater than 1).

4.4 Re-assessing mobility with pre-migration information

How much does the picture of immigrants inter-generational mobility change once we use information from the origin (rather than destination) country? To answer this question, I calculate (and compare) mobility under both information sets.

Mobility is calculated as the difference in probability of obtaining the baccalauréat conditional on social background. For ease of presentation, I group $S_{P,0}$ and $S_{P,2}$ into binary variables. I thus calculate the following expression:

$$\frac{\mathbb{P}(S_{C,2} = \text{High}|S_{P,0} = \text{High}) - \mathbb{P}(S_{C,2} = \text{High}|S_{P,0} = \text{Low})}{\mathbb{P}(S_{C,2} = \text{High}|S_{P,2} = \text{High}) - \mathbb{P}(S_{C,2} = \text{High}|S_{P,2} = \text{Low})} \tag{2}$$

The numerator calculates mobility based on origin country characteristics, the denominator based on the situation in the destination country. If the ratio is below one, mobility measured with pre-migration characteristics is lower than estimated with destination country information. In table 6, I report these numbers, using father's occupation before migration for $S_{P,0}$ and in table 7 using grandfather occupation. In both cases $S_{P,2}$ is measured with father's current occupation.

The ratios are below one, suggesting that immigrants' intergenerational mobility is usually over estimated. The magnitude of the difference is between 12 and 17% respectively for the using father's pre-migration occupation or grandfathers' occupation.

Conditioning on $S_{P,0}$ and $S_{P,2}$ is a great simplification as part of the differences in (2) are explained by origin countries, gender and age. To address this concern, I estimate the same ratio but do not use raw conditional probabilities. Instead, I estimate a probit of $\mathbb{P}(S_{C,2} = \text{High})$ where the explanatory variables are age, gender and origin country. I calculate the difference between the actual and the predicted values of having obtained the baccalauréat and calculate

(2) with the conditional expectation of the 'residuals' instead of the conditional probability of $S_{C,2}$.

Results are reported in tables A12 and A13 respectively when $S_{P,0}$ is measured with father's occupation before migration and grandfathers' occupation. The results are stronger and the difference between the picture that emerges from using information from the origin country sharper. Indeed, intergenerational mobility appears to be respectively 37 and 26% lower.

5 A multi-generational model of immigrants' human capital

5.1 Modelling framework

To explain these findings, I develop a two periods model following Borjas (2015). In the first period, parents arrive in the destination country and work. They cannot transfer the totality of their HC and lose a share $\delta \in (0,1)$ of their raw level. In this period, children are young and not yet on the labor market. Parents face two investment decisions: how much to invest into accumulating back HC for themselves and how much into developing the HC of their children. The costs of these investments is that they eat part of their HC in the first period. π is the share invested in adapting, i.e. *parents investment* and θ the share invested in developing children's HC, i.e. *children investments*. Parents are benevolent, the destiny of their children matters for them as much as their own.

In the second period, both children and parents are on the labor market. Parents have accumulated extra HC since the first period, at rate g, meaning they enjoy $(1+g)$ of what they had in the first period. For simplicity, children start with the HC level of their parents, in the sense that there is full transmission[16]. Their HC between the two periods has grown at rate m.

For parents and children, the technology of HC production is driven by the initial level of HC. If our fictional construction worker wanted to take an exam to have his previous background recognized, he would learn and work with the productivity of a white collar not that of a construction worker. When he invests in his children, for instance by helping them with homework, he also has the productivity of a white collar.

Parents start to accumulate from $(1-\delta)K$. In the destination country, the father is a construction worker and his situation will improve (or not) from that point. On the other hand, I

[16]For ease of presentation, I also assume that parents value the second period as much as the first, i.e. have a discount factor of one.

assume that children start from the initial level of their family. The downgrading that happened in the labor market did not affect the situation inside the household.

Put together, this gives the two following equations of investment's productivity, for *parents* and *children investment*:

$$g(1-\delta)K = (\pi K)^{\alpha} K^{\beta}$$
$$mK = (\theta K)^{\xi} K^{\lambda}$$

Since the investments are made by parents in the destination country in the period following their arrival, HC level are investments taken away from the available HC is $(1-\delta)K$.

The maximization program for parents in the first period is[17]:

$$\max_{\pi,\theta} \left[(1-\delta)K - (\theta+\pi)\,(1-\delta)\,K\right] + \left[(1+g)(1-\delta)K + (1+m)K\right] \tag{3}$$

In the first period, the father enjoys the monetary equivalents of being a construction worker $(1-\delta)K$ minus investments $(\theta+\pi)\,(1-\delta)\,K$. Since only parents work in the first period, they do not benefit from the marketable HC of their children. In the second period, both have accumulated HC at their respective rates and from their respective starting points. The solutions are (I focus on the case where K=1):

$$\pi^* = \left(\frac{1}{1-\delta}\right)^{\frac{1}{1-\alpha}} \alpha^{\frac{1}{1-\alpha}}$$
$$\theta^* = \left(\frac{1}{1-\delta}\right)^{\frac{1}{1-\xi}} \xi^{\frac{1}{1-\xi}}$$

The change in situation for the parents between the two periods and the change between the parents and the children are measured by

$$(1-\delta)(1+g^*) - (1-\delta) = g^*(1-\delta) \tag{4}$$

$$(1+m^*) - (1+g^*)(1-\delta) = (m^*-g^*) + \delta(1+g^*) \tag{5}$$

[17] I assume an interior solution under the conditions that π, θ, α and ξ are inferior to one to avoid having zero or negative marketable capital in the first period.

The very low level of occupational mobility for the parents gives an indication that $g^*(1-\delta) \approx 0$. If I focus on the extreme case where $g^* = 0$, equation 5 shows the importance of including the loss of HC when measuring the mobility between parents and children. If δ is assumed to be zero, the ability of second generation to accumulate HC is overestimated. One crucial dimension of the dynamics that is not fully explained by the evidence presented so far is the term $(m^* - g^*)$.

The rate of HC accumulation for both parents and children can be rewritten as a function of the level of investments and a technological parameter. In particular, by plugging π^* and θ^* into the formulas for g and m, one obtains $\xi m^* = (1-\delta)\theta^*$ and $\alpha g^* = \pi^*$. If we assume that the investment decisions are the same, $\theta^* = \pi^*$, $m^* > g^*$ can be rewritten as $\alpha > \xi$[18]. This means that $(m^* - g^*)$ is driven by a difference in investment productivity. It pays more to invest in children's HC than in parents'.

5.2 Testing empirically the model's predictions

Investments' pay off is small for the first generation. I look at the returns to experience in the host country as an indirect investment from first generation immigrants, I estimate

$$\ln(y_i) = \alpha + \beta_1 \mathrm{exp}_{\mathrm{FR}} + \beta_2 \mathrm{exp}^2_{\mathrm{FR}} + \beta_3 \mathrm{exp}_{\mathrm{before}} + \beta_4 \mathrm{exp}^2_{\mathrm{before}} + \gamma X_i + \varepsilon_i \tag{6}$$

from Friedberg (2000) where $\mathrm{exp}_{\mathrm{FR}}$ refers to the number of years in France, $\mathrm{exp}_{\mathrm{before}}$ the difference between the age at migration and 18 and y the wage. I estimate this regression on males and report the results in the upper part (panels A and B) of table 8. I include origin country fixed effects in column 2 and control for education level in all specifications. To give an idea of how small/large the returns are, I report in column 3 the rates of return to experience[19] for native males.

In the baseline specification, the returns are negative for immigrants (although they are marginally significant). It is important to note that it is based on a sample of immigrants who have adult children. For instance, when re-estimated on all immigrants (panel B), the coefficient is positive, 1% return which is very close to the initial result by Friedberg (2000). What always hold (and what is important here) is that returns are much lower than they are for natives

[18] $m^* = \frac{1-\delta}{\xi} > \frac{1}{\alpha} = g^* \rightarrow \alpha > \xi$ Recall that if a number between 0 and 1 is elevated to a power itself between 0 and 1, a larger value of the exponent means a lower productivity, i.e. $\pi^\xi > \pi^\alpha$ if $\xi, \alpha, \pi \in (0,1)$ and $\alpha > \xi$.

[19] Measured by age-18.

(column 3)[20].

Investments for second generation immigrants are more productive. For each potential outcome used to show resurgence and every measure of investment available in *sample second*, I regress the outcome on the investment interacted with a dummy for immigrants and natives (and a series of individual controls, namely age, gender, country of origin and occupation of the father). This gives the following equation to estimate:

$$y_i = \alpha + \sum_{j=0}^{1} \sum_{l=0}^{1} \beta_{j,l} 1 \left\{ \text{Native}_i = j, \text{Investment}_i = l \right\} + \gamma X_i + \varepsilon_i \tag{7}$$

This allows to test whether the returns to investment are positive but also if they are lower, larger or similar than for natives. Results are reported in table 9. Each sub-table focuses on a type of investment, each column on an outcome.

Returns to parental investment for immigrants are particularly high for educational outcomes. They are either indistinguishable or if anything higher than for natives. The last rows of each panel report the p-value of the F-test associated with the null of equality between the two coefficients. It fails to reject the null (at the 5% level) in 11 out of the 12 regressions. In the remaining case, coefficients for immigrants are larger than for natives. Returns are sometimes negative for natives (although individual coefficients are not significant) which probably captures reverse causality.

Tables 8 and 9 should be interpreted with caution since they are purely descriptive results. However, they both indicate that immigrant parents face problems adapting their HC but not developing that of their children.

6 Conclusion

In this paper, I look at the long-term assimilation of immigrants and see how it is influenced by the socio-economic status pre-migration/in the home country. Using unique French survey data, I reconstitute family histories over three generations, partially in the origin country, partially in France. I find evidence of a U-shape pattern; immigrants were heterogeneous before migrating,

[20] The estimates for immigrants' return to experience is also close to the results of Borjas (1995). The differential between natives and immigrants is thus driven by the high return for natives. This result holds when natives' experience is measured by age-25 and when regressions are weighted.

this heterogeneity is greatly reduced with the first generation and reappears with the second generation. I benchmark immigrants' intergenerational mobility with that of natives using machine learning techniques. I show that immigrants are more sensitive to the background of their parents (at least the part acquired in the origin country) and grandparents than are natives. Using origin (rather than destination) country information changes the picture of immigrants intergenerational mobility; it appears much lower. I explain these facts with a model where parents cannot fully transfer their human capital between labour markets but transmit it across generations.

There are two main takeaways from this paper which are interesting to the general discussion on the success of second generation immigrants. First, when assessing the achievements of second generation immigrants, it is critical to remember that the relevant starting point for this population may not be the situation in the destination country. Social mobility is indeed lower for immigrants when one takes a long-term perspective. The second takeaway is that there is a dichotomy between what happens in the labour market and what happens in the household following migration. Immigrants may be downgraded in the labour market, they still are able to transmit the benefits of their previous socio-economic status to their children.

One of the limitation of this paper is that it does not allow to distinguish between the different elements of human capital. In particular, no information is available on the transmission of behaviours (patience, grit ...) and aspiration (in particular in education). If data collection was to include specific questions about this aspect of parenting, looking at this channel would be an interesting venue for future research.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson, "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

_ , _ , **Elisa Jacome, and Santiago Perez**, "Intergenerational Mobility of Immigrants in the US over Two Centuries," NBER Working Papers 26408, National Bureau of Economic Research, Inc October 2019.

Aydemir, Abdurrahman, Wen-Hao Chen, and Miles Corak, "Intergenerational Earnings Mobility among the Children of Canadian Immigrants," *The Review of Economics and Statistics*, May 2009, *91* (2), 377–397.

Becker, Gary S and Nigel Tomes, "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, July 1986, *4* (3), S1–39.

Biau, Gerard and Erwan Scornet, "A random forest guided tour," *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, June 2016, *25* (2), 197–227.

Black, Sandra E. and Paul J. Devereux, *Recent Developments in Intergenerational Mobility*, Vol. 4 of Handbook of Labor Economics, Elsevier,

Borjas, George J, "The Intergenerational Mobility of Immigrants," *Journal of Labor Economics*, January 1993, *11* (1), 113–35.

_ , "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?," *Journal of Labor Economics*, April 1995, *13* (2), 201–245.

Borjas, Georges J., "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again," *Journal of Human Capital*, 2015, *9* (5), 483–517.

Braun, Sebastian Till and Jan Stuhler, "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany," *Economic Journal*, March 2018, *128* (609), 576–611.

Breiman, Leo, "Random Forests," *Machine Learning*, 2001, *45* (1), 5–32.

Caponi, Vincenzo, "Intergenerational Transmission Of Abilities And Self?Selection Of Mexican Immigrants," *International Economic Review*, 05 2011, *52* (2), 523–547.

Catron, Peter, "The Melting-Pot Problem? The Persistence and Convergence of Premigration Socioeconomic Status During the Age of Mass Migration," *Social Forces*, 2020.

Chan, Tak Wing and Vikki Bolivier, "The Grandparents Effect in Social Mobility: Evidence from British Cohort Studies," *American Sociological Review*, 2013, *92* (1), 662–679.

Clark, Gregory, *The Son Also Rises: Surnames and the History of Social Mobility* number 10181. In 'Economics Books.', Princeton University Press, 2014.

Cohen-Goldner, Sarit and Zvi Eckstein, "Labor Mobility Of Immigrants: Training, Experience, Language, And Opportunities," *International Economic Review*, 08 2008, *49* (3), 837–872.

Eckstein, Zvi and Yoram Weiss, "On The Wage Growth of Immigrants: Israel, 1990-2000," *Journal of the European Economic Association*, June 2004, *2* (4), 665–695.

Friedberg, Rachel M, "You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital," *Journal of Labor Economics*, April 2000, *18* (2), 221–251.

Ichou, Mathieu, "Who They Were There: Immigrants' Educational Selectivity and Their Children's Educational Attainment," *European Sociological Review*, 2014, *30* (6), 750 – 765.

Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *American Economic Review*, 2013, *103* (4), 1109–1137.

Mare, Robert, "A Multigenerational View of Inequality," *Demography*, February 2011, *48* (1), 1–23.

Olivetti, Claudia and M. Daniele Paserman, "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940," *American Economic Review*, August 2015, *105* (8), 2695–2724.

Ward, Zachary, "The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration," *American Economic Journal: Applied Economics*, 2020.

Xie, Yu and Alexandra Killewald, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850: Comment," *American Economic Review*, 2013, *103*, 2003–2020.

Table 1: Descriptive statistics - First Generation Immigrants (Parents) - Immigrants with children

	All		Europe		North Africa		Africa		Asia	
					Gender					
Women	1,796	57	648	56	456	59	264	59	350	52
Men	1,379	43	500	44	322	41	181	41	318	48
					Year of Birth					
25th percentile	1953		1952		1952		1955		1954	
Median	1957		1957		1957		1960		1959	
75th percentile	1962		1961		1961		1964		1965	
					Year of Migration					
25th percentile	1971		1966		1971		1979		1976	
Median	1978		1971		1977		1984		1980	
75th percentile	1984		1978		1984		1989		1985	
					Education Level					
< Primary School	965	30	341	30	287	37	69	16	247	37
Primary School	1,133	36	537	47	250	32	121	27	207	31
Secondary School	341	11	122	11	43	6	64	14	88	13
Higher Education	501	16	140	12	87	11	110	25	92	14
Not available	235	7	8	1	111	14	81	18	34	5
					Occupation Pre-Mig					
Self-Employed Agricultural	97	3	41	4	16	2	12	3	27	4
Self-Employed Non-Agricultural	111	3	19	2	22	3	26	6	36	5
High Managerial	104	3	36	3	13	2	15	3	20	3
Supervisory Occupations	168	5	55	5	31	4	32	7	26	4
Lower Services	256	8	107	9	38	5	50	11	49	7
Lower Technical	423	13	203	18	83	11	35	8	91	14
Not available	2,016	63	687	60	575	74	275	62	419	63
					Education level father					
< Primary School	1,930	61	712	62	604	78	227	51	354	53
Primary School	599	19	237	21	105	13	92	21	135	20
Secondary School	149	5	41	4	16	2	25	6	44	7
Higher Education	250	8	79	7	16	2	52	12	66	10
Not available	247	8	79	7	37	5	49	11	69	10
					Occupation Father					
Self-Employed Agricultural	448	14	112	10	105	13	90	20	121	18
Self-Employed Non-Agricultural	478	15	85	7	124	16	83	19	154	23
High Managerial	215	7	59	5	21	3	43	10	59	9
Supervisory Occupations	193	6	56	5	39	5	49	11	40	6
Lower Services	277	9	49	4	63	8	71	16	78	12
Lower Technical	1,441	45	763	66	390	50	84	19	184	28
Not available	123	4	24	2	36	5	25	6	32	5

Note : The Observations are parents of second generation immigrants followed in sample first. I report the gender of the parent being interviewed, year of birth, year of migration, education level and occupation before migration. I also report the occupation and education level of their father, i.e. the grandfather of second generation immigrants. The entry 965 should be read as follows; among first generation immigrants, 965 had a father who completed less than primary school. The first two columns report raw number and percentages for all origins, the subsequent ones break down the information by region (of the world) of origin.

Table 2: Transition matrix from occupation before and after migration - First Generation Immigrants (Parents) - (Male) Immigrants with children

	S-E Non-Agri	High Managerial	Supervisory Occupations	Lower Services	Lower Technical	Low Occupations	High Occupations
				Percentages			
S-E Agri	0	0	0.04	0.04	**0.93**	**0.96**	0.04
	(0)	(0)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
S-E Non-Agri	0.09	0.06	0.04	0.11	**0.70**	**0.91**	0.09
	(0.04)	(0.03)	(0.03)	(0.04)	(0.06)	(0.04)	(0.04)
High Managerial	0.04	**0.53**	0.07	0.09	**0.27**	**0.40**	0.60
	(0.03)	(0.07)	(0.04)	(0.04)	(0.07)	(0.07)	(0.07)
Supervisory Occ.	0.01	0.08	**0.37**	0.17	**0.37**	**0.55**	0.45
	(0.01)	(0.03)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)
Lower Services	0	0.03	0.05	0.25	**0.66**	**0.92**	0.08
	(0)	(0.02)	(0.03)	(0.06)	(0.06)	(0.04)	(0.04)
Lower Technical	0	0.01	0.04	0.07	**0.88**	**0.95**	0.05
	(0)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
				Raw numbers			
S-E Agri	0	0	2	2	52	54	2
S-E Non-Agri	5	3	2	6	38	49	5
High Managerial	2	24	3	4	12	18	27
Supervisory Occ.	1	6	26	12	26	39	32
Lower Services	0	2	3	15	39	54	5
Lower Technical	0	3	11	20	260	280	14

Note : This table reports the transition between occupation before migration in the origin country (rows) and occupation immediately following migration in France (columns). The upper panel reports conditional probabilities, the lower one raw numbers. The last two columns merge occupations into two categories, "high" for "high managerial" and "supervisory occupations" and low, for the rest. The Observations are fathers of (adult) second generation immigrants followed in sample first.

Table 3: Transition matrix from first to current job - First Generation Immigrants (Parents) - (Male) Immigrants with children

	S-E Agri	S-E Non-Agri	High Managerial	Supervisory Occupations	Lower Services	Lower Technical	Low Occupations	High Occupations
				Percentages				
S-E Agri	0	0	1	0	0	0	0	1
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
S-E Non-Agri	0	**0.76**	0	0.07	0.03	**0.14**	**0.93**	0.07
	(0)	(0.08)	(0)	(0.05)	(0.03)	(0.06)	(0.05)	(0.05)
High Managerial	0	0	**0.84**	**0.10**	0.03	0.03	0.03	**0.97**
	(0)	(0)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Supervisory Occ.	0	0.08	**0.19**	**0.60**	0.04	0.08	0.20	**0.80**
	(0)	(0.02)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
Lower Services	0.01	0.13	0.11	0.13	**0.31**	**0.32**	**0.73**	0.27
	(0.01)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Lower Technical	0.01	0.15	0.02	0.11	0.07	**0.64**	**0.86**	0.14
	(0)	(0.01)	(0)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
				Raw numbers				
S-E Agri	0	0	1	0	0	0	0	1
S-E Non-Agri	0	22	0	2	1	4	28	2
High Managerial	0	0	78	9	3	3	3	90
Supervisory Occ.	0	11	25	79	5	11	26	105
Lower Services	1	19	16	19	45	47	108	40
Lower Technical	9	138	20	97	64	574	796	126

Note : This table reports the transition between first and last occupations in France (columns). The upper panel reports conditional probabilities, the lower one raw numbers. The last two columns merge occupations into two categories, "high" for "high managerial" and "supervisory occupations" and low, for the rest. The Observations are fathers of (adult) second generation immigrants followed in sample first. The sample is limited to immigrants who reported working before migrating.

Table 4: Importance of pre-migration background - Second-generation immigrants (children) - Sample first

	Panel A: Father's pre-migration occupation	
	Father's pre-migration Occupation	Father's current Occupation
S-E Non-Agri	0.18**	0.17
	(0.08)	(0.16)
High Managerial	0.29***	0.35**
	(0.08)	(0.16)
Supervisory Occ.	0.15*	0.28*
	(0.08)	(0.16)
Lower Services	0.14*	0.16
	(0.08)	(0.16)
Lower Technical	0.05	0.15
	(0.06)	(0.15)
F-test	3.98	2.73
N Obs	1,140	
R squared	0.07	
Mean	0.58	
	Panel B: Grandfather's occupation	
	Grandfather's Occupation	Father's current Occupation
S-E Non-Agri	0.16***	0.23
	(0.04)	(0.15)
High Managerial	0.22***	0.49***
	(0.05)	(0.14)
Supervisory Occ.	0.16***	0.37**
	(0.06)	(0.15)
Lower Services	0.08	0.21
	(0.06)	(0.15)
Lower Technical	-0.01	0.22
	(0.04)	(0.14)
F-test	6.85	12.83
N Obs	1,562	
R squared	0.09	
Mean	0.58	

Note : This table reports the estimates of equation 1. Panel A measures $S_{P,0}$ with father's pre-migration occupation while column B use grandfather's occupation. In both panels, $S_{P,2}$ is measured with father's occupation. Control variables include age, and gender. Standard errors are clustered at the family level. For each regression, I report a F-test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$. The observations are (adult) children of first generation immigrants, whose father reported having a job before migration (panel A), whose grandfather stayed in the origin country or migrated after 60 y.o. (panel B).

Table 5: How important is family background - Natives vs Immigrants (children) - Random Forest - Sample first

	Immigrants Country FE	Immigrants no Country FE	Natives
With parents and grandparents information			
Occupation gd father, father	0.98	0.98	0.89
Occupation gd father, mother	0.96	0.92	0.81
Education father	0.98	0.98	0.73
Education mother	0.9	0.85	0.79
Education gd father	0.77	0.81	0.58
Education gd mother	0.73	0.75	0.58
With grandparents information			
Occupation gd father, father	0.96	0.96	0.91
Occupation gd father, mother	0.96	0.93	0.81
Education gd father	0.84	0.87	0.64
Education gd mother	0.81	0.97	0.63

Note : This table reports the importance of each variable, as calculated by random forests (mean decrease impurity criteria). Variable importance is normalized by the importance of father's occupation (value of 1). In the first two columns, the outcome to be predicted is whether immigrant children obtained the baccalauréat and in the last column, this outcome is to be predicted for natives. All forests include age and gender in their algorithms with the variables listed in the rows. The middle column includes country of origin dummies in the algorithms. The lower panel restricts the sample to immigrants whose grandfather did not migrate to France or migrated after the age of 60 years old. In the upper panel, the sample is further restricted to those whose parent completed their education in the origin country.

Table 6: Intergeneration Mobility (parents - children) with pre and post migration characteristics - Father information - Sample first

	Education High (1)	Education Low (2)	(1)-(2)	N High	N Low
Pre-migration criteria	72.77	54.16	18.61	213	962
Post-migration criteria	74.88	53.73	21.15	211	964
Ratio			0.88		

Note : The first two columns calculate the probability of obtaining the baccalauréat for children of immigrants whose father had a "high" and "low" occupation respectively. The third column takes the difference. The last two columns report the number of children immigrants falling in each category. The first row uses father's occupation before migration and the second father's current occupation.

Table 7: Intergeneration Mobility (parents - children) with pre and post migration characteristics - Grandfather information - Sample first

	Education High (1)	Education Low (2)	(1)-(2)	N High	N Low
Pre-migration criteria	75.67	52.69	22.98	526	2101
Post-migration criteria	79.17	51.40	27.77	557	2070
Ratio			0.83		

Note : The first two columns calculate the probability of obtaining the baccalauréat for children of immigrants from a "high" and "low" background respectively. The third column takes the difference. The last two columns report the number of children immigrants falling in each category. The first row uses grandfather's occupation, the second father's occupation.

Table 8: Returns to experience and Language classes - First generation immigrants (parents) - Sample first

	Immigrants		Natives
	Panel A: With children - Returns to experience		
Experience	-0.015*	-0.018*	0.081**
	(0.011)	(0.011)	(0.035)
Experience squared	0.000***	0.000***	-0.001
	(0.000)	(0.000)	(0.001)
N Obs	851	851	323
Control Education	YES	YES	YES
Country FE	NO	YES	NO
	Panel B: With and without children - Returns to experience		
Experience	0.010***	0.010***	0.045***
	(0.003)	(0.003)	(0.004)
Experience squared	0.000***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)
N Obs	2,566	2,566	1,213
Control Education	YES	YES	YES
Country FE	NO	YES	NO

Note: Panels A and B report the estimates of 6. The coefficients reported refer to experience in France (calculated as the number of years since arrival) and its square. Certain specifications include dummies for country of origin. The third column estimates a similar specification for natives where potential experience is calculated as age minus 18. Panel A looks at immigrants and natives with (adult) children, panel B at all immigrants and natives.

Table 9: Returns on Children investment - Natives and Immigrants - Sample second

	D-O	BAC	HE	D-O	BAC	HE	D-O	BAC	HE	D-O	BAC	HE
	Help for Homework			Private Classes			Room alone			Schooling Strategy		
Immigrants	0.029*	0.024	0.032	0.014	-0.003	0.066**	0.073***	0.079***	0.071***	0.041***	0.084***	0.137***
	(0.016)	(0.025)	(0.037)	(0.014)	(0.021)	(0.033)	(0.017)	(0.022)	(0.026)	(0.016)	(0.027)	(0.045)
Natives	0.017	0.064**	-0.043	-0.022	0.019	-0.027	0.037	0.109***	0.027	0.021	0.129***	0.046
	(0.022)	(0.032)	(0.035)	(0.022)	(0.030)	(0.031)	(0.025)	(0.031)	(0.033)	(0.022)	(0.031)	(0.034)
N Obs	6,021	6,021	4,693	6,058	6,058	4,749	6,120	6,120	4,786	6,120	6,120	4,786
N Immigrants	2,447	2,447	1,119	2,415	2,415	1,106	2,462	2,462	1,128	2,462	2,462	1,128
p-value $\mathbb{H}_0$ equality	0.659	0.297	0.102	0.101	0.496	0.017	0.094	0.279	0.162	0.398	0.230	0.081

Note: The table presents the results of equation 7. The outcomes are not dropping out "D-O", obtaining the baccalaureat "BAC" and a higher education degree "H-E". Estimation is performed on a sample of second generation immigrants and natives. The rows refered to as Immigrant and Natives are the coefficients associated with investing and being an immigrant or a native. The four investments are sending children to a school outside the neighborhood "School", helping with homework "Homework", providing a room where children can study alone "Room" and pay for private classes "Private Classes". The baseline category is being a non investing immigrant. Regressions control for age, gender, country of origin and occupation of the father. In each subtable the row p-value H_0 tests the hypothesis of equality between the two coefficients reported.

A1 Appendix

Figure A1: Who is who? (sample first)

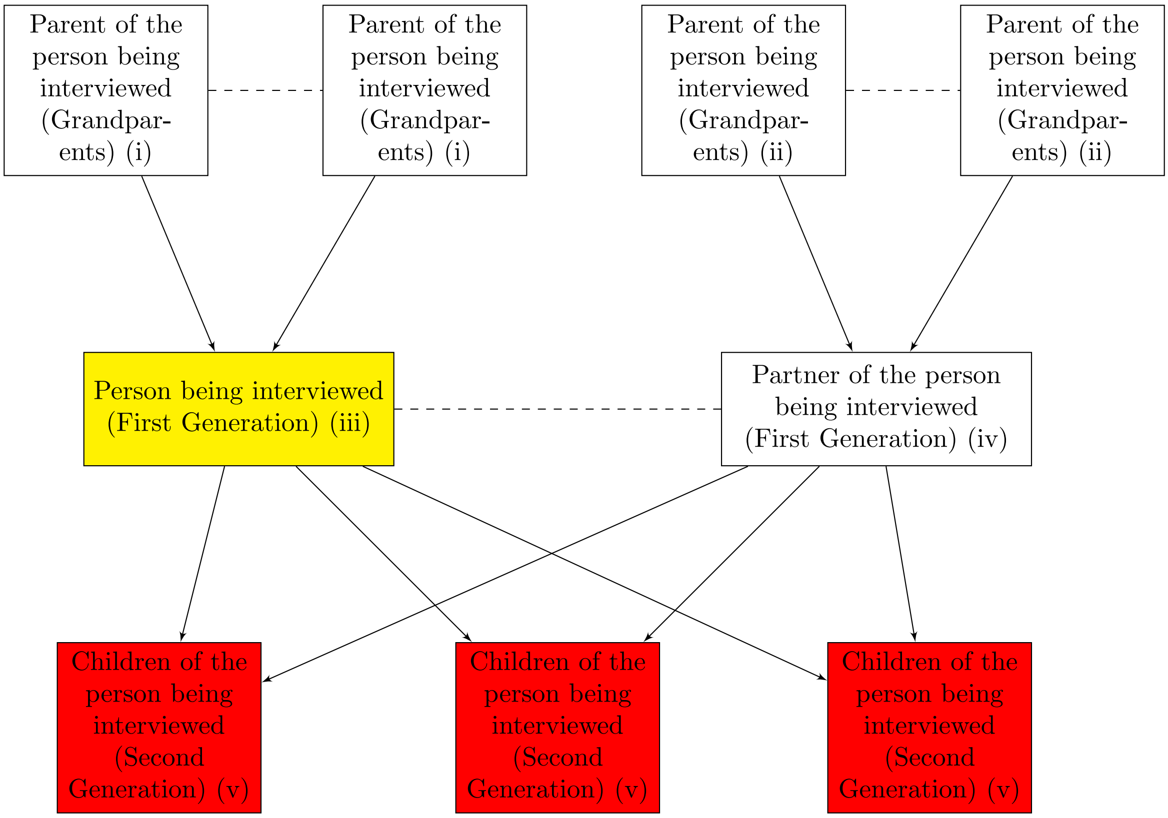

Figure A1: The person being interviewed in TeO is in yellow, the population of interest in red.

Subject	Criteria	Justification
To be part of the sample	(v) must have both parents, i.e. (iii) and (iv), immigrants	The reason why I choose to look at families with both parents immigrants (as opposed to those with one or two parents) is to make sure that I capture the influence of the pre-migration SES. By focusing on children of mixed couples (one immigrant and one native), the influence of the pre-migration status is mixed with the influence of the socio-economic background of the parent who is a native. It can then be that the relevant background for the children is the one of the native parent.
	(v) must have arrived before age 10	To make sure that they have been socialized in the environment characterized by $S_{p,2}$.
	(i) must have arrived after age 60 or not have emigrated to France	Since TeO is a cross section, the variable indicating the occupation of the grandfather in 2008 could be his occupation in France if he had moved. To prevent this case, I keep in the sample families whose grandfather never moved to France or moved after age 60. A caveat is that I can only restrict the sample on the basis of the location (or time of migration) of one of the grandfathers (the father of the parent of the person being interviewed)
Status in France	The National Statistical Agency (INSEE) has a one digit nomenclature of professional occupation with six main categories: Self Employed Agricultural, Self Employed Non Agricultural, High Managerial, Supervisory Occupations, Lower Services and Lower Technical. Are considered "high" status, the individuals whose occupation is classified as High Managerial or Supervisory Occupations, the rest is classified as "low" status.	
	Based on the occupation of the father (iii) if he is a man	
Status before migration	Parents Definition occupation of (iii) or education level of (iii)	
	Grandparents Definition occupation of (i) (or (ii)), or education level of (i)	
Outcomes	Baccalauréat, being 18 or above	18 is a sensitive age to have finished high school.

Figure A2: Who is who? (sample second)

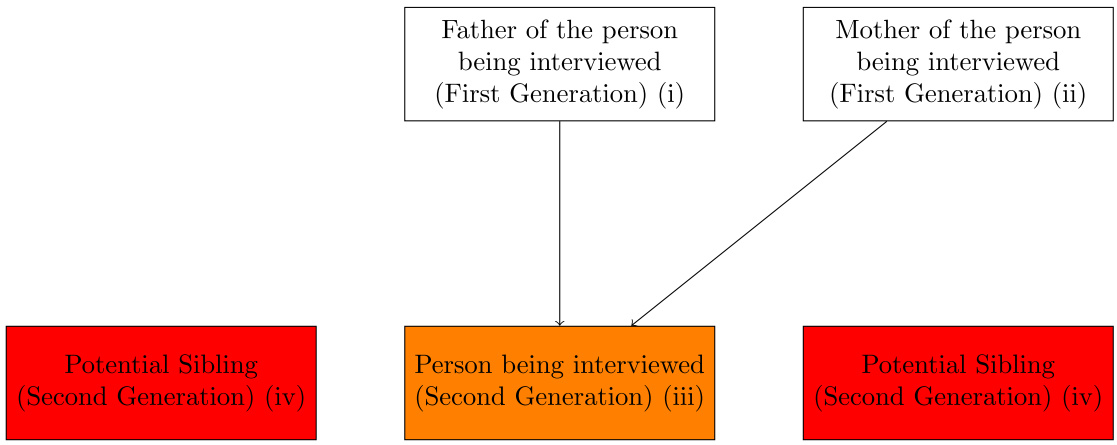

Figure A2: The person being interviewed in TeO is in yellow, the population of interest in red. Since the person interviewed is also from the population of interest, it is orange.

Category of data	Criteria	Justification
To be part of the sample	(iii) must have both parents	The reason why I choose to look at families with both parents immigrants (as opposed to those with one or two parents) is to make sure that I capture the influence of the pre-migration SES. By focusing on children of mixed couples (one immigrant and one native), the influence of the pre-migration status is mixed with the influence of the socio-economic background of the parent who is a native. It can then be that the relevant background for the children is the one of the native parent.
	(i) and (ii) have arrived after age 18	To make sure that parents received (or not) their primary education in their country of origin, I restrict the sample to parents who arrived after age 18
Status in France	The National Statistical Agency (INSEE) has a one digit nomenclature of professional occupation with six main categories: Self Employed Agricultural, Self Employed Non Agricultural, High Managerial, Supervisory Occupations, Lower Services and Lower Technical. Are considered "high" status, the individuals whose occupation is classified as High Managerial or Supervisory Occupations, the rest is classified as "low" status.	
Status before migration	(i) finished secondary education	
Outcomes	Dropout and Baccalauréat, being 18 or above	18 is a sensitive age to have finished high school.
	Higher Education, being 25 or above	25 is a sensitive age to have finished university.
Information on the additional survey	I create several dummy variables if the child has been in a school different than the one of their neighborhood where he was supposed to go, if children report that either their father or their mother helped them sometimes or often with their homework, if children had a room of their own to study and if parents paid for private classes to their children	

Table A1: From one survey to two samples

		Part of the Survey		
		Personal History	Current Situation	Children
Sampled as	first	grandparents 0G parents pre mig 1st G	parents France 1st G	children France 2nd G
	second	parents pre mig 1st G	parents France 1st G children France 2nd G	grandchildren France 3rd G

Note: In each cell is described the population targeted by looking at the sample in the survey (rows) and the questions asked to him/her (columns). In the upper table, I report the population, in the lower one the type of information collected.

Table A2: Descriptive statistics - First Generation Immigrants (Parents) - All sample

	All		Europe		North Africa		Africa		Asia	
					Gender					
Women	4,492	53	1,323	56	1,079	52	902	55	957	49
Men	3,964	47	1,036	44	1,004	48	742	45	994	51
					Year of Birth					
25th percentile	1958		1954		1958		1961		1960	
Median	1966		1962		1966		1969		1969	
75th percentile	1975		1969		1975		1977		1976	
					Year of Migration					
25th percentile	1977		1969		1976		1985		1979	
Median	1988		1980		1987		1993		1986	
75th percentile	1999		1997		1999		2001		1997	
					Education Level					
< Primary School	2,023	24	514	22	631	30	288	18	549	28
Primary School	2,563	30	866	37	617	30	450	27	578	30
Secondary School	1,266	15	349	15	224	11	281	17	329	17
Higher Education	2,240	26	616	26	462	22	474	29	447	23
Not Available	364	4	14	1	149	7	151	9	48	2
					Occupation Pre-Mig					
Self Employed Agricultural	158	2	51	2	29	1	27	2	49	3
Self Employed Non-Agricultural	374	4	65	3	90	4	121	7	84	4
High Managerial	383	5	154	7	57	3	50	3	57	3
Supervisory Occupations	564	7	191	8	115	6	104	6	97	5
Lower Services	795	9	258	11	143	7	191	12	150	8
Lower Technical	949	11	360	15	212	10	126	8	228	12
Not Available	5,233	62	1,280	54	1,437	69	1,025	62	1,286	66
					Education level father					
< Primary School	4,130	49	1,095	46	1413	68	638	39	900	46
Primary School	1,775	21	554	23	345	17	325	20	474	24
Secondary Education	556	7	147	6	80	4	127	8	143	7
Higher Education	1,174	14	355	15	119	6	302	18	240	12
Not available	821	10	208	9	126	6	252	15	194	10
					Occupation Father					
Self Employed Agricultural	835	10	162	7	166	8	215	13	255	13
Self Employed Non-Agricultural	1,289	15	215	9	320	15	278	17	391	20
High Managerial	892	11	262	11	110	5	243	15	163	8
Supervisory Occupations	744	9	196	8	142	7	216	13	141	7
Lower Services	847	10	134	6	193	9	274	17	201	10
Lower Technical	3,470	41	1,324	56	1,070	51	300	18	715	37
Not Available	379	4	66	3	82	4	118	7	85	4

Note : The Observations are first generation immigrants followed in sample first. I report the gender of the parent being interviewed, year of birth, year of migration, education level and occupation before migration. I also report the occupation and education level of their father, i.e. the grandfather of second generation immigrants. The entry 2,023 should be read as follows; among first generation immigrants, 2,023 had a father who completed less than primary school. The first two columns report raw number and percentages for all origins, the subsequent ones break down the information by region (of the world) of origin.

Table A3: Descriptive statistics - First Generation Immigrants (Parents) - Sample second

	Fathers	Mothers
	Year of birth	
25th percentile	1930	1933
Median	1938	1941
75th percentile	1945	1959
	Year of arrival	
25th percentile	1962	1962
Median	1968	1969
75th percentile	1975	1976
	Education level	
< Primary School	1,520	1,710
Primary School	484	431
Secondary School	130	134
Higher Education	176	131
Not available	249	153

Note : The Observations are parents of second generation immigrants followed in sample second. I report for fathers (left column) and mothers (right column), the year of birth, the year of arrival in France and the education level. The entry 131 should be read as follows; among second generation immigrants, 131 have a mother with a higher education degree.

Table A4: How much does mobility change with year of arrival and age? - First generation immigrants (parents) - Immigrants with children

	Outcome: Binary Variable for Low Occupation			
	Panel A: Between pre-migration and arrival			
Higher Occ. pre-migration	-0.45***	-0.46***	-0.45***	-0.36***
	(0.03)	(0.03)	(0.03)	(0.04)
N Obs	579	579	579	579
R squared	0.26	0.29	0.36	0.46
	Panel B: Since arrival			
Low Occ. at arrival	0.72***	0.72***	0.72***	0.68***
	(0.03)	(0.03)	(0.03)	(0.03)
N Obs	1,325	1,323	1,323	1,323
R squared	0.36	0.38	0.41	0.44
Year of birth	NO	YES	YES	YES
Age at arrival	NO	YES	NO	YES
Year of arrival	NO	NO	YES	YES
Country of origin	NO	NO	NO	YES

Note : Panel A reports the results of a regression where the dependent variable is a dummy for having a "low" occupation immediately following migration. The main regressor is a dummy for having a "high" occupation in the origin country. The first column does not include any control, the second includes year of birth fixed effects and age at arrival. The third column includes year of birth and year at arrival fixed effects. The fourth column includes all the controls mentioned and origin country fixed effects. The observations are fathers of (adult) second generation immigrants followed in sample first, who reported having a job before migrating. Panel B reports the same regressions but the dependent variable is having a low occupation at the time of the interview and the main regressor is having a low occupation at arrival. The sample is the same as for panel A but also includes immigrants who did not work before migration.

Table A5: Comparing first-generation immigrants (parents) and natives mobility - Between first and current job

Outcome: Binary Variable for Low Occupation		
Low Occ. - Imm	0.63***	0.63***
	(0.05)	(0.05)
Low Occ. - Nat	0.46***	0.46***
	(0.05)	(0.05)
Year of birth	NO	YES
N Obs	1,705	1,705
R squared	0.32	0.33
F test	46.88	44.81

Note : This table reports the results of a regression where the dependent variable is having a "low" occupation at the time of the interview and the main regressors are the interaction of immigration status and a dummy for "low" first occupation in France. The first column does not include year of birth fixed effects, while the second does. Observations are fathers (immigrants and natives) of adult children. The last row reports the value of the F-test of equality between immigrants and natives.

Table A6: Importance of pre-migration background - Second-generation immigrants (children) - Sample first - Sample weight

	Panel A: Father's pre-migration occupation	
	Father's pre-migration Occupation	Father's current Occupation
S-E Non-Agri	0.16*	0.13
	(0.09)	(0.15)
High Managerial	0.32***	0.36**
	(0.10)	(0.15)
Supervisory Occ.	0.19*	0.31**
	(0.10)	(0.16)
Lower Services	0.17*	0.19
	(0.09)	(0.16)
Lower Technical	0.06	0.16
	(0.07)	(0.14)
F-test	3.08	2.85
R squared	0.07	
Mean	0.58	
	Panel B: Grandfather's occupation	
	Grandfather's Occupation	Father's current Occupation
S-E Non-Agri	0.15***	0.23*
	(0.05)	(0.14)
High Managerial	0.23***	0.50***
	(0.06)	(0.1)
Supervisory Occ.	0.17***	0.37**
	(0.06)	(0.15)
Lower Services	0.10	0.21
	(0.08)	(0.15)
Lower Technical	-0.05	0.24*
	(0.04)	(0.14)
F-test	6.61	9.14
R squared	0.09	
Mean	0.58	

Note : This table reports the estimates of equation 1. Panel A measures $S_{P,0}$ with father's pre-migration occupation while column B use grandfather's occupation. In both panels, $S_{P,2}$ is measured with father's occupation. Control variables include age, and gender. Standard errors are clustered at the family level. Observations are weighted (frequency weights provided by TeO). For each regression, I report a F-test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$. The observations are (adult) children of first generation immigrants, whose father reported having a job before migration (panel A), whose grandfather stayed in the origin country or migrated after 60 y.o. (panel B).

Table A7: Importance of pre-migration background - Second-generation immigrants (children) - Sample first - Origin Country FE

	Panel A: Father's pre-migration occupation	
	Father's pre-migration Occupation	Father's current Occupation
S-E Non-Agri	0.17**	0.08
	(0.08)	(0.16)
High Managerial	0.17***	0.26
	(0.07)	(0.17)
Supervisory Occ.	0.06	0.27
	(0.08)	(0.17)
Lower Services	0.04	0.14
	(0.07)	(0.17)
Lower Technical	0.05	0.08
	(0.06)	(0.15)
F-test	1.60	2.50
p-value	0.16	0.03
N Obs	1,140	
R Squared	0.16	
Mean	0.58	

	Panel B: Grandfather's occupation	
	Grandfather's Occupation	Father's current Occupation
S-E Non-Agri	0.11**	0.27*
	(0.05)	(0.16)
High Managerial	0.14***	0.51***
	(0.05)	(0.15)
Supervisory Occ.	0.11	0.41***
	(0.07)	(0.16)
Lower Services	0.05	0.27*
	(0.06)	(0.16)
Lower Technical	-0.03	0.24
	(0.04)	(0.15)
F-test	3.12	10.13
N Obs	1,562	
R squared	0.14	
Mean	0.58	

Note : This table reports the estimates of equation 1. Panel A measures $S_{P,0}$ with father's pre-migration occupation while column B use grandfather's occupation. In both panels, $S_{P,2}$ is measured with father's occupation. Control variables include age, gender and country of origin dummies. Standard errors are clustered at the family level. For each regression, I report a F-test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$. The observations are (adult) children of first generation immigrants, whose father reported having a job before migration (panel A), whose grandfather stayed in the origin country or migrated after 60 y.o. (panel B).

Table A8: Importance of pre-migration background Second generation immigrants (children) - Sample Second

	Dropping-Out	Baccalauréat	Higher Education
Primary School	0.08***	0.16***	0.20***
	(0.02)	(0.03)	(0.04)
Secondary School	0.13***	0.26***	0.42***
	(0.02)	(0.04)	(0.09)
Higher Education	0.14***	0.24***	0.38***
	(0.02)	(0.05)	(0.09)
F-test parents occupation in France	0.74	2.67	0.84
F-test education parents	20.09	22.83	16.85
Nb Obs	2,252	2,252	2,252
Mean	0.86	0.53	0.3

Note : This table reports the estimates of equation 1. Each column uses a definition of $S_{C,2}$ (column label). Control variables include age, gender and dummy for father's occupation. Standard errors are robust to heteroskedasticity. For each regression, I report a F-test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$. The observations are (adult) second generation immigrants, whose both parents emigrated after 25 y.o.

Table A9: Importance of pre-migration background - Second-generation immigrants (children) - Sample first - Varying $S_{P,2}$ (1)

	Panel A: Father's pre-migration occupation	
	Father's pre-migration Occupation	Father's first Occupation
S-E Non-Agri	0.18**	
	(0.09)	
High Managerial	0.27***	0.40***
	(0.09)	(0.16)
Supervisory Occ.	0.16*	0.29*
	(0.09)	(0.17)
Lower Services	0.11	0.21
	(0.08)	(0.17)
Lower Technical	0.02	0.24
	(0.07)	(0.16)
F-test	3.20	2.68
N Obs	1,098	
R squared	0.06	
Mean	0.58	
	Panel B: Grandfather's occupation	
	Grandfather's Occupation	Father's first Occupation
S-E Non-Agri	0.16***	
	(0.04)	
High Managerial	0.21***	0.22*
	(0.05)	(0.13)
Supervisory Occ.	0.17***	0.13
	(0.06)	(0.13)
Lower Services	0.08	0.00
	(0.06)	(0.13)
Lower Technical	-0.03	-0.02
	(0.04)	(0.12)
F-test	7.29	6.40
N Obs	1,609	
R squared	0.08	
Mean	0.58	

Note : This table reports the estimates of equation 1. Panel A measures $S_{P,0}$ with father's pre-migration occupation while column B use grandfather's occupation. In both panels, $S_{P,2}$ is measured with father's first occupation. Control variables include age, and gender. Standard errors are clustered at the family level. For each regression, I report a F-test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$. The observations are (adult) children of first generation immigrants, whose father reported having a job before migration (panel A), whose grandfather stayed in the origin country or migrated after 60 y.o. (panel B).

Table A10: Importance of pre-migration background - Second-generation immigrants (children) - Sample first - Varying $S_{P,2}$ (2)

	Panel A: Father's pre-migration occupation
S-E Non-Agri	0.14
	(0.09)
High Managerial	0.34***
	(0.08)
Supervisory Occ.	0.11*
	(0.09)
Lower Services	0.21***
	(0.08)
Lower Technical	0.07
	(0.07)
F-test	6.14
N Obs	899
R squared	0.06
Mean	0.57
	Panel B: Grandfather's occupation
S-E Non-Agri	0.15***
	(0.05)
High Managerial	0.27***
	(0.06)
Supervisory Occ.	0.16**
	(0.08)
Lower Services	0.12*
	(0.07)
Lower Technical	-0.01
	(0.04)
F-test	7.19
N Obs	1,176
R squared	0.07
Mean	0.57

Note : This table reports the estimates of equation 1. Panel A measures $S_{P,0}$ with father's pre-migration occupation while column B use grandfather's occupation. Control variables include age, gender and total resources available to the household (as a measure of $S_{P,2}$). Standard errors are clustered at the family level. For each regression, I report a F-test of $\mathbb{H}_0 : S_{P,0} = 0$ and $\mathbb{H}_0 : S_{P,2} = 0$. The observations are (adult) children of first generation immigrants, whose father reported having a job before migration (panel A), whose grandfather stayed in the origin country or migrated after 60 y.o. (panel B).

Table A11: How important is family background - Natives vs Immigrants (children) - Random Forest - Sample second

	Immigrants Country FE	Immigrants no Country FE	Natives
	Dropping Out		
Education father	1.07	1.18	0.77
Education mother	1.04	1.13	0.66
	Obtaining Baccalauréat		
Education father	1.06	1.15	0.82
Education mother	1.01	1.06	0.76
	Obtaining a Higher Education degree		
Education father	1.09	1	0.79
Education mother	1.13	1.04	0.81

Note : This table reports the importance of each variable, as calculated by random forests (mean decrease impurity criteria). Variable importance is normalized by the importance of father's occupation (value of 1). In the first panel, the outcome to be predicted is whether immigrant children dropped out and in the last column, this outcome is to be predicted for natives. In the middle panel the outcome is obtaining the baccalauréat and in the last one, a higher education degree. All forests include age and gender in their algorithms with the variables listed in the rows. The middle column includes country of origin dummies in the algorithms. The immigrant sample is composed of second generations whose parents immigrated after age 25.

Table A12: Intergeneration Mobility (parents - children) with pre and post migration characteristics - Father information - Residuals - Sample first

	Education High (1)	Education Low (2)	(1)-(2)	N High	N Low
Pre-migration criteria	7.46	-1.19	8.65	213	961
Post-migration criteria	11.67	-2.09	13.76	211	963
Ratio			0.63		

Note : The first two columns calculate the means of the 'residuals' of the probability of obtaining the baccalauréat for children of immigrants whose father had a "high" and "low" occupation respectively. Residuals are the difference between actual success and a probit model which includes age, gender and country of origin fixed effects. The third column takes the difference. The last two columns report the number of children immigrants falling in each category. The first row uses father's occupation before migration and the second father's current occupation.

Table A13: Intergeneration Mobility (parents - children) with pre and post migration characteristics - Grandfather information - Residuals - Sample first

	Education High (1)	Education Low (2)	(1)-(2)	N High	N Low
Pre-migration criteria	10.65	-2.03	12.68	523	2094
Post-migration criteria	13.95	-3.10	17.05	553	2064
Ratio			0.74		

Note : The first two columns calculate the means of the 'residuals' of the probability of obtaining the baccalauréat for children of immigrants from a "high" and "low" background respectively. Residuals are the difference between actual success and a probit model which includes age, gender and country of origin fixed effects. The third column takes the difference. The last two columns report the number of children immigrants falling in each category. The first row uses grandfather's occupation, the second father's occupation.